В. С. ПУГАЧЕВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА



В. С. ПУГАЧЕВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Допущено Министерством высшего и среднего специального образования СССР в качестве учебного пособия для студентов высших технических учебных заведений



МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1979

22.17
П 88
УДК 519

Теория вероятностей и математическая статистика. Пугачев В. С. — М.: Наука. Главная редакция физико-математической литературы, 1979.

В книге изложены основы теории вероятностей и математической статистики. В первых пяти главах дается достаточно строгое изложение основ теории вероятностей в рамках конечномерных случайных величин на основе традиционных курсов математического анализа и линейной алгебры. В последующих пяти главах изложены основы математической статистики: точечное и интервальное оценивание параметров распределений, плотностей и функций распределения, общая теория оценок, метод стохастических аппроксимаций, методы построения статистических моделей.

Книга предназначена для студентов и аспирантов факультетов прикладной математики вузов и для инженеров.

Владимир Семенович Пугачев

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

М., 1979 г., 496 стр. с илл.

Редактор *Е. Ю. Ходан.*

Технический редактор *С. Я. Шкляр.* Корректоры *Т. С. Плетнева, Л. С. Сомова.*

ИБ № 2293

Сдано в набор 11.07.79. Подписано к печати 11.11.79. Т-18648. Бумага $84\times108^1/_{32}$, тип. № 3. Литературная гарнитура. Высокая печать. Условн. печ. л. 26,04. Уч.-изд. л. 29,24. Тираж 55000 экз. (1-й завод 1—30000 экз.). Заказ № 413. Цена книги 1 р. 20 к.

Издательство «Наука» Главная редакция физико-математической литературы 117071, Москва, В-71, Ленинский проспект, 15

Ордена Октябрьской Революции и ордена Трудового Красного Знамени Первая Образцовая типография имени А. А. Жданова Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. Москва, М-54, Валовая, 28

Отпечатано в тип. № 4 изд-ва «Наука», Новосибирск 77, Станиславского, 25 Заказ № 662.

П $\frac{20203-176}{053(02)-79}$ 14-79. 1702060000

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Книга написана на основе курсов лекций, читанных автором на протяжении ряда лет студентам факультета прикладной математики Московского авиационного института им. С. Орджоникидзе (МАИ).

Книга рассчитана на студентов и аспирантов факультетов прикладной математики университетов и высших технических учебных заведений. Однако она может быть полезной и для инженеров и научных работников, которым приходится пользоваться статистическими методами в прикладных исследованиях, а также для математиков, специализирующихся в области теории вероятностей и математической статистики.

Книга ориентирована в первую очередь на прикладников. Это определило ее структуру и характер. В ней дано достаточно строгое изложение основ теории вероятностей и математической статистики в рамках конечномерных случайных величин без применения теории меры и функционального анализа.

За основу построения теории вероятностей принята система аксиом А. Н. Колмогорова. Однако аксиомы вводятся лишь после изучения свойств частот событий и подхода к понятию вероятности как абстрактному понятию, отражающему экспериментально наблюдаемую закономерность в поведении частот событий — их устойчивость. В результате такого подхода аксиомы теории вероятностей вводятся как естественное распространение свойств частот событий на вероятности.

При изучении случайных величин почти везде, особенно в задачах математической статистики, рассматриваются сразу векторные случайные величины. Это усиливает прикладную направленность книги, так как в большей части задач практики приходится иметь дело

с многомерными случайными векторами (конечными множествами скалярных случайных величин).

Для того чтобы приблизить излагаемые методы к непосредственному практическому применению с использованием ЭВМ, в книге даны ссылки на соответствующие стандартные программы вычислений в «Сборнике научных программ на Фортране» фирмы IBM [94].

Первоначально книга была задумана только как учебное пособие по одноименному курсу, однако затем возникла необходимость включить в книгу дополнительный материал по математической статистике, необходимый для выполнения студентами курсовых и дипломных работ, а также обеспечения студенческих научно-исследовательских работ. В результате книга, кроме основ теории вероятностей, содержит краткое изложение всех основных разделов математической статистики (применительно к конечномерным случайным величинам).

Кроме традиционных вопросов точечного и интервального оценивания и общей теории оценок, в книге изложены метод стохастических аппроксимаций, многомерный регрессионный анализ, дисперсионный анализ, факторный анализ, теория оценивания неизвестных параметров в стохастических разностных уравнениях, основы теории распознавания и проверки гипотез, элементы общей статистической теории решений, основы метода статистического моделирования.

В первой главе изучаются основные свойства частот событий, дается частотный подход к понятию вероятности и рассматриваются случаи, когда вероятности событий можно вычислять непосредственно из соображений равновозможности (равновероятности) различных исходов опыта. После этого дается понятие элементарного события, формулируются основные аксиомы теории вероятностей, вводятся понятия вероятностного пространства, распределения вероятностей, условной вероятности, зависимости и независимости событий и выводятся основные формулы, непосредственно вытекающие из аксиом, в том числе формулы, определяющие биномиальное и полиномиальное распределения. Затем выводится распределение Пуассона.

Во второй главе рассматриваются случайные величины и их распределения, изучаются основные характеристики распределений конечномерных случайных величин — плотность и функция распределения. Показывается, что плот-

ность как обобщенная функция, содержащая линейную комбинацию δ-функций, существует для всех трех типов случайных величин, встречающихся в задачах практики,— непрерывных, дискретных и непрерывно-дискретных. Приводится пример случайной величины, не имеющей плотности такого типа.

В третьей главе изучаются числовые характеристики случайных величин. Сначала дается определение математического ожидания и изучаются основные свойства математических ожиданий. Потом дается определение моментов второго порядка и изучаются их свойства. После этого определяются моменты любых порядков для действительных случайных величин. Кроме моментов, для действительных скалярных случайных величин даются понятия медианы и квантилей. Глава заканчивается изучением одномерного нормального распределения.

В четвертой главе изучаются распределения и условные распределения проекций случайного вектора. Выводятся формулы для определения плотности проекции случайного вектора и ее условной плотности при данном значении проекции случайного вектора на дополнительное подпространство по данной плотности случайного вектора. Даются понятия зависимости и независимости случайных величин. Определяются условные моменты. Изучаются многомерное нормальное распределение и характеристические функции случайных величин.

В пятой главе излагаются методы нахождения распределений функции случайных величин по данным распределениям величин-аргументов. Рассматриваются общий метод определения функций распределения функций случайных величин, два метода определения плотностей — метод сравнения элементов вероятности и метод δ-функций — и метод определения характеристических функций. Дается доказательство предельной теоремы для сумм независимых случайных величин в случае одинаково распределенных слагаемых. В качестве примеров применения общих методов приводится вывод основных распределений, встречающихся в математической статистике.

В шестой главе сначала дается постановка основной задачи математической статистики — задачи оценивания неизвестных вероятностей событий, распределений случайных величин и их параметров. Затем рассматриваются основные виды сходимости последовательностей случайных величин, даются общие определения, касающиеся оценок

и доверительных областей, и излагаются основные методы нахождения доверительных областей для неизвестных параметров. После этого изучаются частота как оценка вероятности события и оценки моментов, получаемые с помощью выборочных средних. Глава заканчивается изложением основных методов проверки гипотез о параметрах распределений.

В седьмой главе излагается общая теория оценок параметров распределений и основные методы нахождения оценок — метод максимума правдоподобия и метод моментов. Изучается рекуррентное оценивание корня уравнения регрессии и точки экстремума регрессии методом стохастических аппроксимаций.

В восьмой главе излагаются основные методы оценивания плотностей и функций распределения случайных величин и методы приближенного аналитического представления распределений. Излагаются методы проверки гипотез о распределениях по критериям К. Пирсона, А. Н. Колмогорова и Н. В. Смирнова и рассматривается оценивание параметров распределений методом минимума χ^2. В последнем параграфе дается краткое изложение метода статистического моделирования как метода приближенных вычислений и как метода научного исследования.

В девятой главе изучаются регрессионные статистические модели. Сначала излагается общий метод определения средней квадратической регрессии в заданном классе функций, в частности линейной средней квадратической регрессии. Потом излагаются методы оценивания линейных регрессий (регрессионный анализ) и методы проверки гипотез о регрессиях. Наконец, из общей теории построения линейных регрессионных моделей выводятся основы теории дисперсионного анализа.

В десятой главе изучаются статистические модели других видов. Сначала рассматриваются модели, описываемые разностными уравнениями, в частности модели авторегрессии, излагается метод оценивания последовательностей случайных величин, определяемых разностными уравнениями, и неизвестных параметров в разностных уравнениях и применение этого метода к линейным и нелинейным моделям авторегрессии. Затем изучаются некоторые методы построения факторных моделей (элементы факторного анализа) и модели распознавания. Показывается общность некоторых задач распознавания

и задач проверки гипотез о параметрах распределений. В последнем параграфе дается краткое изложение элементов статистической теории решений (методов построения моделей процессов принятия решений).

В каждой главе книги принята своя нумерация параграфов, пунктов, формул и примеров. При ссылках на параграфы, пункты, формулы и примеры в пределах одной главы указываются только их номера в этой главе. При ссылках на параграфы, пункты, формулы и примеры из других глав перед номером параграфа, пункта, формулы или примера ставится номер соответствующей главы, отделенный от номера параграфа, пункта, формулы или примера точкой. Так, например, § 2, п. 3.4, (72) и пример 5 означают ссылки на § 2, п. 3.4, формулу (72) и пример 5 той же главы, в которой даны эти ссылки; § 5.3, п. 5.2.3, (5.26) и пример 5.17 означают ссылки на § 3, п. 2.3, формулу (26) и пример 17 главы 5.

Ссылки на литературу даны номерами соответствующих литературных источников в приложенном в конце книги списке, заключенными в квадратные скобки. Автор ни в какой мере не претендует на полноту приложенного списка литературы. В нем указаны только те источники, на которые даются ссылки в тексте.

Для удобства читателей формулировки всех основных результатов и предложений выделены курсивом. Начало и конец выводов, доказательств и рассуждений, приводящих к определенным результатам, отмечены зачерненными треугольными указателями ▶ и ◀.

В главах 6—10 дано лишь краткое изложение основных методов современной математической статистики конечномерных случайных величин. Для более глубокого и полного изучения математической статистики можно рекомендовать книги Г. Крамера [45], М. Кендалла и А. Стьюарта [35], С. Уилкса [103], С. Р. Рао [86], Т. Андерсона [1], а также книги по различным разделам математической статистики, на которые даны ссылки в главах 6—10.

Для изучения математических основ теории вероятностей рекомендуем книги М. Лоэва [51], Ж. Невё [64] и П. Л. Хеннекена и А. Тортра [119].

Для справок по математическому анализу рекомендуем книги В. И. Смирнова [96, т. 1, т. 2], Г. М. Фихтенгольца [109], С. М. Никольского [69], по линейной алгебре—книги В. И. Смирнова [96, т. 3, ч. 1], А. И. Маль-

цева [56], Л. Н. Головиной [18], по теории матриц — книги Ф. Гантмахера [16], П. Ланкастера [48], М. Маркуса и Х. Минца [57].

Последний параграф главы 8 («Метод статистического моделирования») и главы 9, 10 написаны при активном участии И. Н. Синицына, который помог также отредактировать всю рукопись. Без его помощи книга, вероятно, еще не скоро была бы написана. Считаю своим приятным долгом выразить И. Н. Синицыну глубокую признательность за его неоценимую помощь.

Глубокую благодарность приношу также Н. И. Андрееву и Н. М. Сотскому за ценные замечания и дискуссии, способствовавшие значительному улучшению рукописи, И. В. Синицыной, взявшей на себя труд неоднократной перепечатки различных вариантов рукописи, Н. С. Беловой, О. В. Тимохиной, А. С. Пиунихину, И. Д. Силуяновой и Н. Т. Чулюкановой за помощь в оформлении рукописи, С. Я. Виленкину за консультации по вычислительным аспектам изложенных в книге методов и за организацию вычислений на ЭВМ для ряда примеров.

В. С. Пугачев

Москва
Апрель 1978 г.

Глава 1

ВЕРОЯТНОСТИ СОБЫТИЙ

§ 1. Случайные явления

1.1. Примеры случайных явлений. Человек в своей практической деятельности на каждом шагу встречается со случайными явлениями. Без них не протекает ни один процесс. Простейшим примером случайных явлений служат ошибки измерений. Мы знаем, что абсолютно точных измерений не существует, и чем точнее измерительный прибор, тем это более заметно. Измеряя один и тот же предмет, например взвешивая его на аналитических весах много раз, мы всегда получаем близкие, но различные результаты. Это объясняется тем, что результат каждого измерения содержит случайную ошибку и результаты разных измерений содержат различные ошибки. Предвидеть, какова будет ошибка данного конкретного измерения, или даже определить ее после измерения принципиально невозможно. Производя экспериментальное исследование какого-либо явления и систематизируя результаты в виде графических зависимостей, мы убеждаемся в том, что экспериментальные точки, если их достаточно много, никогда не ложатся на одну кривую, а всегда заполняют некоторую полосу, т. е. имеет место случайный разброс экспериментальных точек. Этот разброс объясняется как ошибками измерений, так и действием других случайностей.

Вторым примером случайных явлений может служить рассеивание снарядов. Снаряды никогда не ложатся в одну и ту же точку, даже тогда, когда прицеливание производится по одной точке. Казалось бы, условия одинаковы для всех выстрелов. Однако снаряды летят по разным траекториям и попадают в различные точки. Предвидеть заранее, в какую именно точку попадет данный снаряд, принципиально невозможно потому, что мы не можем абсолютно точно знать параметры состояния атмосферы во всех точках траектории, по которой летит

снаряд, а от этих параметров зависят аэродинамические силы и их моменты, действующие на снаряд.

В качестве третьего примера случайных явлений укажем отказы различных технических устройств. Как бы совершенна ни была современная техника, иногда происходят отказы тех или иных приборов. Отказ прибора — случайное явление. Невозможно заранее предвидеть, откажет он или нет, и если откажет, то в какой момент времени.

К случайным явлениям относятся также шумы при приеме радиопередач. Так называемый «эфир» всегда насыщен различными электромагнитными излучениями, источниками которых служат электрические разряды в атмосфере, движение атмосферного электричества, работающие электрические устройства, созданные человеком, и т. п. Поэтому, как бы мы ни настраивали приемник на излучение интересующей нас станции, в него всегда попадают посторонние излучения, которые мешают слушать станцию, особенно удаленную. Это проявляется в том, что наряду с принимаемой передачей мы слышим в репродукторе шум и потрескивания. Это хорошо знакомое всем явление тоже представляет собой случайное явление, так как невозможно заранее предсказать, когда и какое постороннее электромагнитное излучение попадет в приемник. Избежать же попадания посторонних излучений в приемник принципиально невозможно, раз он предназначен для приема слабых электромагнитных излучений.

Испытываемая нами иногда при полете на самолете «болтанка» — тоже случайное явление. Она представляет собой случайные колебания самолета, вызываемые случайными порывами ветра в неспокойной атмосфере.

1.2. Природа случайных явлений. Как и всякие явления, случайные явления вызываются вполне определенными причинами. Все явления окружающего нас мира взаимно связаны и влияют одно на другое (закон всеобщей связи явлений). Поэтому каждое наблюдаемое явление связано причинной зависимостью с бесчисленным множеством других явлений и течение его зависит от бесчисленного множества факторов. Проследить все это бесконечное множество связей и определить действие каждой из них принципиально невозможно. Поэтому, изучая то или иное явление, человек ограничивается лишь основными факторами, определяющими его течение,

и пренебрегает огромным количеством второстепенных явлений. Это дает возможность глубже проникнуть в сущность явления, установить его закономерность. Вместе с тем, поступая так, человек обедняет явление, схематизирует его. Иными словами, он заменяет изучаемое явление подходящей упрощенной его моделью. Вледствие этого любой закон науки отражает сущность изучаемого явления, но он всегда значительно беднее, у́же самого явления. Никакой закон не может характеризовать явление всесторонне, во всей полноте и многообразии. Наблюдаемые в реальном явлении отклонения от закономерности, вызываемые совместным действием бесчисленного множества неучтенных факторов, и представляют собой случайные явления.

При экспериментальном изучении какого-либо явления с целью установления его закономерностей приходится наблюдать его многократно в одинаковых условиях. При этом под одинаковыми условиями мы понимаем одинаковые значения всех количественных характеристик контролируемых факторов. Все неконтролируемые факторы будут при этом различными. Вследствие этого действие контролируемых факторов будет практически одинаковым при разных наблюдениях одного и того же явления. В этом как раз и проявляются законы данного явления. Случайные же отклонения от закономерности, вызванные действием неконтролируемых факторов, будут различными при разных наблюдениях, причем предвидеть заранее, какими они будут при данном конкретном наблюдении, принципиально невозможно.

Роль случайностей в разных явлениях различна. В некоторых явлениях случайные отклонения от закономерностей настолько малы, что их можно не учитывать. Однако есть и такие явления, в которых невозможно подметить никаких закономерностей и случайность играет основную роль. Примером такого явления может служить движение малой частицы твердого вещества, взвешенной в жидкости, так называемое *броуновское движение*. Под действием толчков огромного количества движущихся молекул жидкости частица движется совершенно беспорядочно, без всякой видимой закономерности. В подобных явлениях сама случайность является закономерностью.

При многократном наблюдении случайных явлений в них самих можно заметить определенные закономер-

ности. Изучив эти закономерности, человек получает возможность в известной степени управлять случайными явлениями, ограничивать их влияние, предсказывать результаты их действия и даже целенаправленно использовать их в своей практической деятельности. Так, например, можно проектировать измерительные системы, обладающие максимальной доступной точностью, радиоприемные устройства с максимальной помехозащищенностью, обладающие минимальным уровнем шумов, системы управления движением летательных аппаратов, обеспечивающие наибольшую возможную точность навигации или наименьшее действие «болтанки» на летательный аппарат. Можно также проектировать технические системы, обладающие заданной надежностью.

1.3. Массовые случайные явления. Из сказанного ясно, что закономерности случайных явлений могут проявляться только при их многократном наблюдении. Отсюда можно сделать вывод, что изучению поддаются только такие случайные явления, которые можно, по крайней мере принципиально, наблюдать много, практически неограниченное число раз. Такие случайные явления называются *массовыми*. Здесь следует заметить, что для изучения массовых случайных явлений нет никакой необходимости в том, чтобы все их можно было фактически наблюдать много раз. Изучив закономерности простейших случайных явлений и построив на основе этого изучения соответствующую теорию, можно после этого теоретически изучать и более сложные случайные явления, в том числе и такие, которые не поддаются непосредственному наблюдению (но которые принципиально мыслимо наблюдать неограниченное число раз). Так, например, в процессе проектирования космического корабля, предназначенного для одного-единственного полета, можно исследовать надежность всего комплекса средств, обеспечивающих полет, произвести их расчет так, чтобы получить практическую уверенность в том, что все эти средства будут работать безотказно и полет пройдет благополучно. Сила науки в том и состоит, что, будучи основана на небольшом числе простейших положений, взятых из непосредственного наблюдения, она позволяет открывать и предсказывать новые факты теоретическим путем, без непосредственных наблюдений.

1.4. Предмет теории вероятностей. Изучением закономерностей массовых случайных явлений занимается осо-

бая математическая наука — *теория вероятностей*. Методы теории вероятностей, называемые *вероятностными* или *статистическими*, дают возможность производить расчеты, позволяющие делать определенные практические выводы относительно случайных явлений. Как и всякая прикладная наука, теория вероятностей нуждается в исходных экспериментальных данных для расчетов. Раздел теории вероятностей, изучающий методы обработки результатов опытов и получения из них необходимых данных, называется *математической статистикой*.

Теория вероятностей является мощным инструментом исследования, и поэтому она находит большое число самых разнообразных применений в различных областях науки и инженерной практики. Области ее применения непрерывно расширяются. В прошлом веке теория вероятностей получила применение в теории измерений, в теории стрельбы и в физике. В нашем веке она постепенно проникла в аэродинамику и гидродинамику, радиотехнику, теорию управления, динамику полета, теорию связи, строительную механику, теорию механизмов и машин, теорию волнения моря и качки кораблей, метеорологию и во многие другие области знания. Сейчас трудно назвать отрасль науки, которая не пользовалась бы вероятностными методами. В современной теории процессов управления, в теоретической радиотехнике теория вероятностей стала основным инструментом исследований. Вся теория современных сложных систем и процессов управления основана на применении статистических методов. Теория вероятностей служит фундаментом для теории надежности технических систем и для многих других прикладных научных теорий.

Этот процесс непрерывного расширения областей применения теории вероятностей вполне естествен и легко объясняется. Дело в том, что в начале развития каждой отрасли науки человек стремится открыть основные законы этой науки и ему достаточно довольно грубого совпадения результатов расчета с данными опытов. Кроме того, техника эксперимента на начальной стадии несовершенна и не может обеспечить высокую точность измерений. По мере развития науки требования к точности расчетов повышаются, техника эксперимента совершенствуется, и случайные явления, которыми можно было пренебрегать в начале развития данной отрасли науки, начинают играть все более и более значительную роль.

В результате старая теория начинает во многом расходиться с экспериментальными данными и возникает необходимость обратиться к теории вероятностей. Теория вероятностей во всех таких случаях неизменно дает новую теорию, более точно описывающую изучаемые явления и обеспечивающую совпадение результатов теоретических расчетов с экспериментальными данными. Так случилось в начале тридцатых годов с теорией турбулентности в аэродинамике и в сороковых годах с теорией автоматического управления и радиотехникой, а потом и с другими прикладными научными теориями.

Особенность вероятностных методов состоит в том, что они рассматривают исследуемое явление в целом, изучают результаты совокупного действия всех причинных связей, которые невозможно проследить по отдельности.

§ 2. Статистический подход к описанию случайных явлений

2.1. Опыт, событие, случайная величина. Исходным пунктом для построения теории вероятностей, как и любой другой теоретической науки, служат некоторые экспериментальные факты, на основе которых формируются соответствующие абстрактные понятия. Чтобы рассказать об этих фактах, необходимо ввести некоторые термины.

Будем называть *опытом* наблюдение какого-нибудь явления при выполнении некоторого комплекса условий и действий, который должен каждый раз строго выполняться при повторении данного опыта. Наблюдение того же явления при другом комплексе условий и действий будет уже другим опытом.

Результаты опыта можно характеризовать качественно и количественно.

Качественная характеристика опыта состоит в регистрации какого-нибудь факта, т. е. в определении того, обладают результаты опыта каким-либо свойством или нет. Любой такой факт называется *событием*. При этом говорят, что «событие появилось (произошло)» или «событие не появилось (не произошло)» в результате опыта.

Примерами событий могут служить отказ прибора в данном интервале времени, попадание и промах при выстреле, получение m попаданий при n выстрелах.

События обозначаются прописными латинскими буквами, обычно начальными, например A, B, C.

Количественная характеристика опыта состоит в определении значений некоторых величин, полученных в результате опыта. Такие величины, которые могут принимать в результате опыта различные значения, причем до опыта невозможно предвидеть, какими именно они будут, называются *случайными величинами*.

Примерами случайных величин могут служить ошибки и результаты измерений, время безотказной работы прибора или системы, рост и вес наугад выбранного человека, координаты точки попадания (разрыва снаряда) при выстреле, число попаданий при n выстрелах.

Мы будем обозначать случайные величины прописными буквами, преимущественно из конца латинского алфавита, а их конкретные значения — соответствующими малыми буквами. Например, случайные величины будем обозначать X, Y, Z, а их конкретные значения, получаемые в результате опыта,— соответственно x, y, z. Эти значения называются *возможными значениями* или *реализациями* случайных величин X, Y, Z.

С каждой случайной величиной можно связать различные события. Типичным событием, связанным со случайной величиной, является событие, состоящее в том, что эта случайная величина примет в результате опыта какое-нибудь значение, принадлежащее данному множеству, безразлично какое именно. Такое событие коротко называется *попаданием случайной величины в данное множество*.

2.2. Частота события. Естественно сравнивать события по тому, как часто каждое из них появляется при повторении данного опыта. Если при повторении опыта одно событие появляется чаще, чем другое, то говорят, что первое вероятнее второго. Ясно, что для сравнения событий необходимо предположить, что данный опыт можно производить сколько угодно раз. В дальнейшем будем говорить для краткости «производится n опытов» вместо «данный опыт производится n раз».

Частотой события называется отношение числа его появлений к числу всех произведенных опытов.

Таким образом, если при n опытах событие A появилось m раз, то его частота в данной серии опытов равна m/n.

2.3. Условные частоты. В некоторых случаях частоту события приходится определять при дополнительном условии, что произошло некоторое другое событие. Чтобы

определить частоту события A при условии, что произошло событие B, необходимо учитывать не все произведенные опыты, а только те из них, в которых произошло событие B.

Таким образом, если из n произведенных опытов событие B появилось в m опытах, причем в k из этих m опытов появилось и событие A, то частота события A при условии, что произошло событие B, равна k/m. Эта частота, как правило, не совпадает с частотой события A, вычисленной с учетом всех n произведенных опытов.

Частота события A, вычисленная с учетом только тех из произведенных опытов, в которых появилось событие B, называется *условной частотой* события A относительно события B.

2.4. Свойства частот. Событие называется *невозможным* и обозначается $\varnothing$, если оно не может произойти в результате данного опыта. Событие называется *достоверным* и обозначается Ω, если оно обязательно происходит в результате данного опыта, т. е. не может не произойти.

События $A_1, \ldots, A_n$ называются *несовместными* в данном опыте, если в результате этого опыта никакие два из них не могут появиться вместе. Например, попадание и промах при одном выстреле, выпадение 1 очка, 2 очков и 3 очков при одном бросании игральной кости.

Два события, несовместные в одном опыте, могут оказаться совместными в другом опыте. Например, попадание и промах несовместны при одном выстреле. Однако они совместны, если за один опыт считаются два выстрела.

После приведенных определений можно перейти к изучению основных свойств частот событий.

1) Частота любого события представляет собой неотрицательное число, не превосходящее 1, причем частота невозможного события равна 0, а частота достоверного события равна 1.

2) Частота появления одного из несовместных событий, безразлично какого именно, равна сумме их частот. Это следует непосредственно из того, что число появлений сложного события, представляющего собой появление какого-нибудь из несовместных событий, равно сумме чисел появлений этих событий.

3) Частота совместного появления двух событий A и B равна частоте одного из них, умноженной на условную частоту другого. Для доказательства достаточно заметить,

что если при n опытах A появилось m раз, а B — k раз, причем l раз вместе с A, а $k-l$ раз без A, то частота совместного появления A и B равна l/n, частота A равна m/n, а условная частота B относительно A равна l/m.

Очевидно, что если частота события в данной серии опытов равна 0 (или 1), то из этого не следует, что событие невозможно (достоверно). Так, например, если при пяти бросаниях монеты герб не появился ни разу, то из этого не следует, что появление герба невозможно.

2.5. Вероятность события. Замечательный экспериментальный факт—основная закономерность, наблюдаемая в массовых случайных явлениях,— устойчивость частот событий при большом числе опытов. Если при малом числе опытов частота события принимает совершенно случайно различные значения, то при неограниченном увеличении числа опытов она проявляет тенденцию стабилизироваться около некоторого характерного для данного события значения.

Допустим, что некоторый опыт неограниченно повторяется и после каждого опыта вычисляется частота события с учетом всех уже произведенных опытов. При этом обнаруживается, что вначале, когда число произведенных опытов мало, случайный результат каждого опыта существенно изменяет частоту события. Однако по мере возрастания числа опытов влияние результата каждого нового опыта уменьшается. Так, например, результат тысячного опыта изменяет частоту меньше, чем на 0,001. Частота как бы стремится перестать быть случайной и стабилизироваться около некоторого значения.

Устойчивость частот событий дает основание считать, что с каждым событием связано некоторое число—*вероятность* этого события,— около которого стремится стабилизироваться его частота. Так, например, частота появления герба при бросании монеты, очевидно, должна стабилизироваться около 1/2. Следовательно, вероятность появления герба равна 1/2.

Вероятность события A обозначается $P(A)$. Это, конечно, не исключает применения сокращенных обозначений, например $P(A)=p$ и т. п.

Понятие вероятности события является первичным в теории вероятностей и поэтому не нуждается в определении. Оно представляет собой результат абстракции, необходимой для построения любой теории. Отвлекаясь

от сложных и несущественных колебаний частоты при неограниченном повторении опытов и оставляя основную, существенную закономерность, наблюдаемую в данном явлении,— устойчивость частоты,—мы и вводим абстрактное понятие вероятности события.

Вероятность события в данном опыте—его объективная характеристика. Она имеет вполне определенное значение независимо от того, собираемся мы производить опыты или нет.

2.6. Выборочное среднее. Основная задача экспериментального изучения случайных величин состоит в том, чтобы установить, как распределяются экспериментальные точки на числовой оси, на плоскости или в пространстве.

Совокупность значений наблюдаемых величин, полученных в результате опытов, называется *выборкой*.

Получив выборку, необходимо прежде всего определить положение значений случайной величины на числовой прямой и их рассеивание, т. е. размеры занимаемой ими области.

За характеристику положения экспериментальных точек обычно принимают среднее арифметическое значение случайной величины, называемое *выборочным средним*.

Предположим, что случайная величина X приняла в результате n опытов значения $x_1, \ldots, x_n$. Тогда выборочное среднее определяется формулой

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k. \qquad (1)$$

2.7. Выборочные дисперсия и среднее квадратическое отклонение. За экспериментальную характеристику рассеивания значений скалярной случайной величины обычно принимают среднее арифметическое значение квадратов отклонений экспериментальных значений случайной величины от выборочного среднего. Эта характеристика называется *выборочной дисперсией* случайной величины.

Если в результате n опытов случайная величина X приняла значения $x_1, \ldots, x_n$, то ее выборочная дисперсия определяется формулой

$$d_x^* = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2. \qquad (2)$$

Недостатком выборочной дисперсии с практической точки зрения является отсутствие наглядности—она имеет

размерность квадрата случайной величины. Поэтому за практическую характеристику рассеивания значений случайной величины обычно принимают *выборочное среднее квадратическое отклонение*, представляющее собой положительный квадратный корень из выборочной дисперсии, $\sigma_x^* = +\sqrt{d_x^*}$.

Для практических расчетов формуле (2) часто придают более удобную форму. Имея в виду, что на основании (1)

$$\sum_{k=1}^{n} (x_k - \bar{x})^2 = \sum_{k=1}^{n} x_k^2 - 2\bar{x} \sum_{k=1}^{n} x_k + n\bar{x}^2 = \sum_{k=1}^{n} x_k^2 - n\bar{x}^2,$$

можем переписать (2) в виде

$$d_x^* = \sigma_x^{*2} = \frac{1}{n} \sum_{k=1}^{n} x_k^2 - \bar{x}^2. \qquad (3)$$

При этом, чтобы избежать разности больших чисел, в случае необходимости из всех экспериментальных значений $x_1, \ldots, x_n$ вычитают произвольное число, расположенное в середине интервала, занимаемого этими значениями (правило «ложного нуля»).

2.8. Метод наименьших квадратов. При совместном изучении нескольких случайных величин, кроме их выборочных средних и дисперсий, необходимо определить еще какие-то характеристики зависимости между ними.

Для отыскания приближенных зависимостей между изучаемыми экспериментально величинами обычно пользуются методом наименьших квадратов.

Предположим, что требуется найти зависимость между наблюдаемыми величинами x и y (не обязательно случайными). Для этого обычно выбирают подходящую функцию, зависящую от некоторых параметров, $\varphi(x;\ c_1, \ldots, c_N)$, например линейную комбинацию N определенных функций с неопределенными коэффициентами, и подбирают параметры $c_1, \ldots, c_N$ так, чтобы сумма квадратов ошибок приближенной зависимости $y = \varphi(x;\ c_1, \ldots, c_N)$ во всех экспериментальных точках была минимальной:

$$\delta = \sum_{k=1}^{n} [y_k - \varphi(x_k;\ c_1, \ldots, c_N)]^2 = \min.$$

В этом и состоит метод наименьших квадратов.

Для минимизации δ можно пользоваться различными методами отыскания экстремума функции, в зависимости от способа задания функции и сложности ее вычисления. В частности, можно применить обычный метод приравнивания нулю первых производных δ по $c_1, \ldots, c_N$ и решения полученных уравнений относительно $c_1, \ldots, c_N$ с последующим исследованием поведения функции δ в окрестности найденного решения.

2.9. Выборочные ковариация и коэффициент корреляции. Предположим, что случайные величины X и Y приняли в результате n опытов пары значений $x_1, y_1, \ldots \ldots, x_n, y_n$. Чтобы найти подходящую характеристику зависимости между величинами X и Y, подберем методом наименьших квадратов наилучшую линейную зависимость Y от X вида

$$y-\bar{y}=c(x-\bar{x}), \tag{4}$$

где $\bar{x}$ и $\bar{y}$ — выборочные средние величин X и Y соответственно.

Для определения коэффициента c в (4) методом наименьших квадратов приравниваем нулю производную по c суммы квадратов ошибок во всех экспериментальных точках

$$\delta=\sum_{k=1}^{n}[y_k-\bar{y}-c(x_k-\bar{x})]^2.$$

Решив полученное уравнение относительно c, находим в силу (2)

$$c=\frac{\sum\limits_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y})}{\sum\limits_{k=1}^{n}(x_k-\bar{x})^2}=\frac{1}{nd_x^*}\sum_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y}).$$

Величина

$$k_{xy}^*=\frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y}) \tag{5}$$

называется *выборочной ковариацией* величин X и Y. За меру зависимости между X и Y принимают безразмерную величину

$$r_{xy}^*=\frac{k_{xy}^*}{\sigma_x^*\sigma_y^*}=\frac{k_{xy}^*}{\sqrt{d_x^*d_y^*}}, \tag{6}$$

называемую *выборочным коэффициентом корреляции* величин X и Y.

Так как

$$\sum_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y})=\sum_{k=1}^{n}x_ky_k-\bar{x}\sum_{k=1}^{n}y_k-\bar{y}\sum_{k=1}^{n}x_k+n\bar{x}\bar{y}=$$
$$=\sum_{k=1}^{n}x_ky_k-n\bar{x}\bar{y},$$

то формулу (5) можно переписать в виде

$$k_{xy}^{*}=\frac{1}{n}\sum_{k=1}^{n}x_ky_k-\bar{x}\bar{y}. \tag{7}$$

Этой формулой обычно и пользуются на практике. При этом в случае необходимости из координат всех экспериментальных точек вычитают соответствующие координаты какой-нибудь точки, расположенной в середине области, занятой точками, чтобы избежать разностей близких величин.

Пять чисел $\bar{x}$, $\bar{y}$, d_x^*, d_y^*, k_{xy}^* или $\bar{x}$, $\bar{y}$, σ_x^*, σ_y^*, r_{xy}^* служат грубой характеристикой распределения экспериментальных точек на плоскости. Точно так же распределение экспериментальных точек в m-мерном пространстве, полученное в результате совместного наблюдения m случайных величин, можно характеризовать выборочными средними, дисперсиями (средними квадратическими отклонениями) и ковариациями (коэффициентами корреляции) *).

2.10. Гистограмма. Чтобы получить более полное представление о распределении экспериментальных точек, обычно разбивают занятую ими область на интервалы (прямоугольники, параллелепипеды) и вычисляют частоты попадания в эти интервалы (прямоугольники, параллелепипеды). Разделив эти частоты на длины интервалов (площади прямоугольников, объемы параллелепипедов), получают относительные плотности экспериментальных точек в соответствующих частях области, занятой экспериментальными точками. Полученное таким образом рас-

---

*) Для вычисления выборочных средних, дисперсий, ковариаций и коэффициентов корреляции в «Сборнике научных программ на Фортране» [94] служит стандартная программа CORRE.

пределение экспериментальных точек можно изобразить графически, построив на каждом интервале прямоугольник, высота которого равна значению относительной плотности экспериментальных точек в этом интервале (рис. 1). Полученная в результате ступенчатая кривая называется *гистограммой*.

При вычислении гистограммы интервалы (прямоугольники, параллелепипеды) рекомендуется выбирать так,

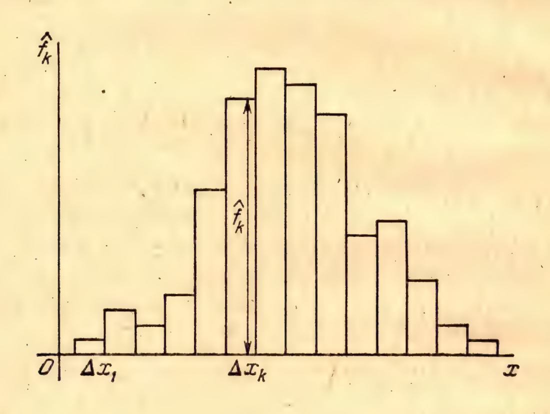


Рис. 1.

чтобы в каждом интервале было по меньшей мере 10 точек (для обеспечения этого можно брать интервалы разной длины). Легко понять, что гистограмму можно построить только при достаточно большом числе экспериментальных точек (для одномерной гистограммы не меньше 100) *).

2.11. Группированная выборка. Если одновременно с построением гистограммы определяются выборочные средние, дисперсии и ковариации, то для упрощения вычислений обычно считают все экспериментальные точки в данном интервале (прямоугольнике) совпадающими с его центром. Полученная таким путем новая выборка называется *группированной выборкой*.

---

*) Для построения одномерной гистограммы в [94] служат программы HIST и TAB1, а для построения двумерной гистограммы — TAB2.

Если пользоваться группированной выборкой, то формулы (1)—(3), (5) и (7) можно записать в виде

$$\bar{x} = \sum_{\nu=1}^{r} \hat{p}_\nu \bar{x}_\nu,$$

$$d_x^{*2} = \sigma_x^{*2} = \sum_{\nu=1}^{r} \hat{p}_\nu (\bar{x}_\nu - \bar{x})^2 = \sum_{\nu=1}^{r} \hat{p}_\nu \bar{x}_\nu^2 - \bar{x}^2, \qquad (8)$$

$$k_{xy}^* = \sum_{\mu,\ \nu=1}^{r} \hat{p}_{\mu\nu} (\bar{x}_\mu - \bar{x})(\bar{y}_\nu - \bar{y}) = \sum_{\mu,\ \nu=1}^{r} \hat{p}_{\mu\nu} \bar{x}_\mu \bar{y}_\nu - \bar{x}\bar{y},$$

где r — число интервалов, содержащих экспериментальные точки, на каждой из осей x, y, $\bar{x}_1, \ldots, \bar{x}_r$, $\bar{y}_1, \ldots, \bar{y}_r$ — координаты центров интервалов на соответствующих осях, $\hat{p}_1, \ldots, \hat{p}_r$ — частоты попадания в интервалы оси x, а $\hat{p}_{\mu\nu}$ ($\mu, \nu = 1, \ldots, r$) — частоты попадания в прямоугольники, на которые разбита плоскость xy *).

Так же как частота события лежит в основе абстрактного понятия вероятности события, рассмотренные экспериментальные характеристики случайных величин лежат в основе соответствующих абстрактных понятий теории вероятностей.

§ 3. Непосредственное определение вероятностей

3.1. Равновозможные исходы опыта. Приведенный в п. 2.4 пример опыта с бросанием монеты показывает, что вероятности некоторых событий легко определяются непосредственно. Рассмотрим общую схему таких опытов. Допустим, что опыт имеет n возможных исходов, так что каждое конкретное его осуществление обязательно заканчивается одним и только одним из этих n исходов, причем нет никаких оснований считать, что при неограниченном повторении опыта какой-нибудь один исход может появляться чаще, чем любой другой. В этом случае вероятность каждого исхода, очевидно, равна $1/n$, так как их частоты должны стабилизироваться при повторении опыта около одного и того же числа, а в сумме они должны дать 1. Иными словами, данный опыт имеет n

---

*) Для вычисления выборочных средних, дисперсий и некоторых других характеристик по группированной одномерной выборке в [94] служит программа MOMEN, использующая для вычисления частот программу TAB1.

равновозможных исходов. В нашем примере с бросанием монеты таких равновозможных исходов два — появление герба и появление цифры, — и вероятность каждого из них равна 1/2.

3.2. Схема случаев. Допустим теперь, что при n равновозможных исходах опыта нас интересует некоторое событие A, связанное с m из этих n исходов так, что при любом из этих m исходов оно обязательно появляется и не может появиться ни при одном из остальных $n-m$ исходов. В таком случае, как легко сообразить, вероятность события A равна отношению m/n. Принято говорить, что в данном опыте имеется n случаев и из них m благоприятствуют событию A. Тогда вероятность события A равна отношению числа случаев, благоприятствующих событию A, к числу всех равновозможных несовместных случаев: $P(A)=m/n$.

Пример 1. Опыт состоит в бросании игральной кости, имеющей форму куба, на каждой грани которого проставлено соответствующее число очков от 1 до 6. В данном опыте имеется 6 случаев: появление одного очка, двух, трех, четырех, пяти и шести. Так как нет никаких оснований предполагать, что, например, два очка могут появляться чаще, чем пять очков; поэтому вероятность появления каждого данного числа очков, например двух, равна 1/6. Рассмотрим теперь событие A — появление четного числа очков, событие B — появление числа очков, кратного трем, и событие C — появление не меньше трех очков. Событию A благоприятствуют три случая — появление двойки, четверки и шестерки; поэтому $P(A)=3/6=1/2$. Событию B благоприятствуют два случая — появление тройки и шестерки; поэтому $P(B)=2/6=1/3$. Событию C благоприятствуют четыре случая — появление тройки, четверки, пятерки и шестерки; поэтому $P(C)=4/6=2/3$.

Пример 2. В урне 10 одинаковых шаров. Из них 3 белых и 7 черных. Шары тщательно перемешивают и после этого из урны вынимают один шар. Найти вероятность появления белого шара.

В данном случае нет никаких оснований предполагать, что при повторении опыта какой-нибудь определенный шар будет появляться чаще, чем любой другой. Например, если шары пронумерованы от 1 до 10 так, что определить номер шара на ощупь невозможно, то нет никаких оснований предполагать, что шар № 1 может появляться при повторении опыта чаще, чем шар № 2, шар № 3 и т. д. Поэтому в данном опыте имеются 10 возможных случаев. Из них 3 благоприятствуют появлению белого шара, а 7 — появлению черного шара. Поэтому вероятность появления белого шара равна 0,3, а вероятность появления черного шара равна 0,7.

3.3. Геометрические вероятности. В некоторых случаях, когда опыт имеет бесконечное (несчетное) множество равновозможных исходов, вероятности некоторых

событий можно определять геометрически как отношения длин отрезков, или площадей, или объемов соответствующих фигур.

Пример 3. Поезда метро идут в данном направлении с интервалом 1 мин. Какова вероятность того, что пассажиру придется ждать поезда не больше 20 с?

В данном случае нет никаких оснований считать какой-нибудь один момент прихода пассажира в интервале между поездами более вероятным, чем любой другой. Интересующему нас событию A — ожиданию меньше 20 с — благоприятствуют все моменты интервала длительностью 1 мин, отстоящие от его конца меньше чем на 20 с. Поэтому вероятность события A можно определить как отношение длительностей двух интервалов времени:

$$P(A) = 20/60 = 1/3.$$

Пример 4. Самолет имеет несколько уязвимых для снарядов данного калибра агрегатов, при попадании в каждый из которых самолет выходит из строя. При попадании в любую другую часть самолета он может продолжать полет. Площадь проекции самолета на плоскость, перпендикулярную траектории снаряда, составляет 100 м^2. Площадь проекции уязвимых частей самолета — 12 м^2. Рассеивание снарядов достаточно велико для того, чтобы можно было считать, что при большом числе попаданий точки попадания распределяются на проекции самолета равномерно. Найти вероятность поражения самолета при попадании в него одного снаряда.

В данном случае имеется бесконечное множество равновозможных исходов опыта — попаданий в различные точки проекции самолета. Допущение о равномерности распределения большого числа попаданий по площади проекции самолета дает основания считать все возможные точки попадания в пределах проекции цели равновозможными. Тогда вероятность поражения самолета при одном попадании, равную вероятности того, что точка попадания придется на проекцию какого-нибудь уязвимого агрегата, можно определить как отношение площади проекции уязвимых агрегатов ко всей площади проекции самолета. В результате вероятность поражения самолета при одном попадании в него будет равна $12/100 = 0{,}12$.

Пример 5. На плоскость с нанесенными на ней параллельными прямыми на расстоянии $2a$ одна от другой случайно бросается прямолинейный отрезок длины $2l$. Найти вероятность того, что отрезок пересечет хотя бы одну из прямых (задача Бюффона об игле).

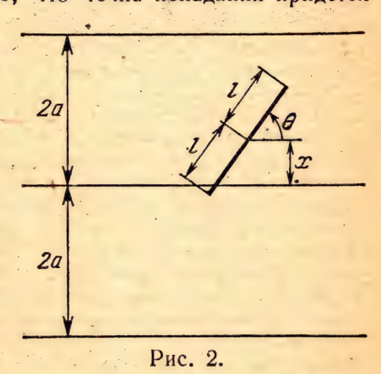


Рис. 2.

Положение отрезка относительно прямых можно характеризовать двумя координатами: расстоянием x его центра от ближайшей прямой, $0 \leqslant x \leqslant a$, и углом θ его наклона к прямым, $0 \leqslant \theta \leqslant \pi/2$ (рис. 2). Очевидно, что отрезок пересекает хотя бы одну из прямых

тогда и только тогда, когда $x \leqslant l \sin \theta$. Если рассматривать θ и x как прямоугольные декартовы координаты точки на плоскости, то область всех возможных пар (θ, x) будет представлять собой прямоугольник со сторонами a и $\pi/2$ (рис. 3). Событие A — выполнение неравенства $x \leqslant l \sin \theta$ — произойдет тогда и только тогда, когда точка (θ, x) попадет в заштрихованную область, ограниченную снизу синусоидой $x = l \sin \theta$ (на рис. 3 показан случай $l < a$). Считая все положения точки (θ, x) в пределах прямоугольника равновероятными, можем вычислить искомую вероятность пересечения отрезка с одной из прямых как отношение площади заштрихованной области к площади прямоугольника. При $l \leqslant a$ это даст $P(A) = 2l/\pi a$. Предлагаем читателю самостоятельно решить эту задачу при $l > a$. В этом случае можно поставить задачу определения вероятностей пересечения двух, трех и т. д. прямых.

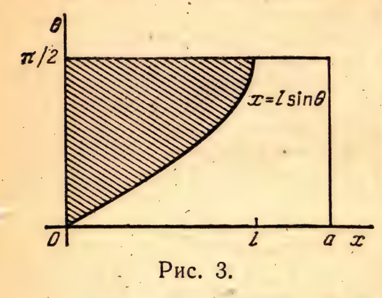


Рис. 3.

3.4. Вычисление условных вероятностей в схеме случаев. Точно так же вычисляются в таких случаях условные вероятности.

П р и м е р 6. В урне 10 шаров — 3 белых и 7 черных. Один шар вынули и отложили в сторону. Он оказался белым. Найти вероятность того, что второй вынутый шар окажется белым.

Пусть событие A — первый вынутый шар оказался белым, B — второй вынутый шар оказался белым. В данном случае нас интересует условная вероятность события B относительно A. Учитывая условие — появилось событие A, т. е. один шар вынут и он оказался белым, — приходим к выводу, что имеется 9 возможных случаев и из них 2 благоприятствуют событию B. Поэтому $P(B \mid A) = 2/9$.

П р и м е р 7. Видоизменим теперь условия предыдущего примера. Предположим, что один шар из урны вынули и, не глядя, положили в другую урну. Найти вероятность того, что второй вынутый шар окажется белым.

Здесь, несмотря на то, что один шар из урны вынут, возможны 10 случаев, так как неизвестно, какой именно шар отсутствует. Из них 3 благоприятствуют появлению белого шара. Следовательно, вероятность появления белого шара остается равной 3/10. Чтобы сделать это совершенно ясным, предположим, что не один, а 9 из 10 шаров переложили, не глядя, в другую урну. В урне остался один шар, но он может быть любым из 10. Поэтому, несмотря на то, что в урне всего один шар, возможных случаев 10, а благоприятствующих появлению белого шара 3.

Таким образом, несмотря на удаление части шаров, речь в данной задаче идет о безусловной вероятности появления белого шара. Объясняется это тем, что нет никакого дополнительного условия для данного опыта, вследствие чего вероятность появления белого шара не изменилась после удаления из урны части шаров. Условием, изменяющим вероятность события, при котором она заменяется условной вероятностью, всегда служит появление некоторого события или

предположение о том, что это событие появилось. Поэтому удаления части шаров недостаточно для того, чтобы можно было говорить об условных вероятностях. Необходимо еще произвести наблюдение и определить, какие именно шары удалены из урны, или сделать предположение о том, какие шары отсутствуют.

Пример 8. В урне 10 шаров — 3 белых и 7 черных. Один шар из урны вынут, он оказался белым. После этого шар вернули в урну и все шары тщательно перемешали. Найти вероятность того, что во второй раз появится белый шар.

В данном случае имеется вполне определенное условие — появилось событие A — первый вынутый шар оказался белым. Поэтому можно говорить об условной вероятности события B — второй вынутый шар оказался белым. Однако число возможных случаев осталось равным 10, а число случаев, благоприятствующих событию B, осталось равным 3. Следовательно, в данном случае

$$P(B \mid A) = P(B) = 3/10.$$

§ 4. Действия над событиями

4.1. Объединение двух событий. *Объединением* или *суммой* двух событий A и B называется сложное событие, состоящее в появлении хотя бы одного из событий A и B. Объединение событий A и B обозначается $A \cup B$. Для несовместных событий A и B применяется также обозначение $A + B$.

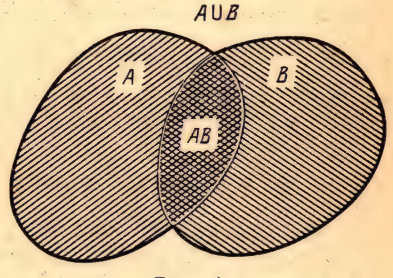


Рис. 4.

Для иллюстрации на рис. 4 показано объединение двух событий A и B для случая, когда событие A представляет собой попадание случайной точки в область, обозначенную буквой A, а событие B — попадание в область, обозначенную буквой B. Событие $A \cup B$ в этом

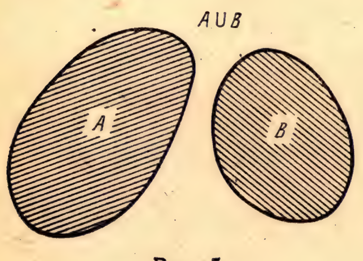


Рис. 5.

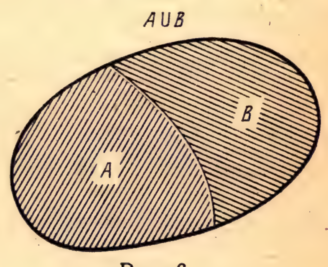


Рис. 6.

случае представляет собой попадание в область, граница которого обведена жирной линией. На рис. 4 показан случай, когда области A и B имеют общую часть (пересекаются), т. е. когда события A и B совместны. На рис. 5 и 6 показаны случаи, когда области A и B не пересекаются, т. е. когда события A и B несовместны.

4.2. Пересечение двух событий. *Пересечением* или *произведением* двух событий A и B называется их совместное появление. Пересечение событий A и B обозначается $A \cap B$ или, короче, AB.

На рис. 4 дана иллюстрация пересечения событий A и B в случае, когда событие A представляет собой попадание в область A, а событие B — попадание в область B. Пересечение AB событий A и B представляет собой попадание в отмеченную двойной штриховкой общую часть областей A и B.

Приведенные определения объединения и пересечения событий легко распространяются на любое число событий.

4.3. Объединение и пересечение любого множества событий. *Объединением* или *суммой* множества событий A_s, $s \in S$, которое обозначается $\bigcup\limits_{s \in S} A_s$ или $\sum\limits_{s \in S} A_s$ (для несовместных событий), называется появление хотя бы одного из событий A_s, $s \in S$.

Пересечением или *произведением* множества событий A_s, $s \in S$, которое обозначается $\bigcap\limits_{s \in S} A_s$, называется совместное появление всех этих событий.

Множество S значений индекса s в этих определениях может быть конечным, счетным или несчетным.

Если, например, событие A_s представляет собой попадание случайной точки в интервал $(s, s+1)$ числовой оси, а S — конечное множество $\{0, 1/2, 1, 3/2, 2\}$, то $\bigcup\limits_{s \in S} A_s$ представляет собой попадание в конечное объединение интервалов

$$(0, 1) \cup (1/2, 3/2) \cup (1, 2) \cup (3/2, 5/2) \cup (2, 3),$$

т. е. в интервал $(0, 3)$. Если S — счетное множество $\{2^{-p}\}_{p=0}^{\infty}$, то $\bigcup\limits_{s \in S} A_s$ представляет собой попадание в счетное объединение интервалов $\bigcup\limits_{p=0}^{\infty} (2^{-p}, 2^{-p}+1)$, т. е. в интервал $(0, 2)$. Наконец, если S — интервал $(0, 1/2)$, то

$\bigcup_{s \in S} A_s$ представляет собой попадание в несчетное объединение интервалов $(s, s+1)$ при всех $s \in (0, 1/2)$, т. е. в интервал $(0, 3/2)$.

4.4. Свойства объединений и пересечений. Операции объединения и пересечения событий обладают рядом свойств, аналогичных свойствам сложения и умножения чисел. Например, *объединение и пересечение событий коммутативны*:

$$A \cup B = B \cup A, \qquad AB = BA.$$

Объединение и пересечение событий ассоциативны:

$$(A \cup B) \cup C = A \cup (B \cup C) = (A \cup C) \cup B = A \cup B \cup C,$$
$$(AB)\,C = A\,(BC) = (AC)\,B = ABC.$$

Наконец, *объединение и пересечение событий дистрибутивны*:

$$(A \cup B)\,C = AC \cup BC.$$

Все эти свойства непосредственно следуют из определений объединения и пересечения событий. Так, $(A \cup B)\,C$ представляет собой совместное появление события C с событием A, или с событием B, или с A и B вместе. Событие $AC \cup BC$ тоже состоит в появлении или C вместе с A, или C вместе с B, или C вместе с AB.

Однако не все законы сложения и умножения чисел справедливы для объединения и пересечения событий. Так, например, события $A \cup A$ и AA, очевидно, совпадают с A. Следовательно, $A \cup A = AA = A$ для любого события A.

4.5. Противоположные события. Для дальнейшего изучения действий над событиями необходимо ввести еще понятие противоположных событий.

Событием, *противоположным* событию A, называется непоявление A, которое обозначается $\bar{A}$.

Легко видеть, что событие A противоположно событию $\bar{A}$:

$$\bar{\bar{A}} = A.$$

Примерами противоположных событий могут служить попадание и промах при выстреле, отказ прибора в данном интервале времени и его исправная работа в том же интервале времени.

Очевидно, что противоположные события несовместны, а их объединение представляет собой достоверное событие:

$$A\overline{A}=\varnothing, \quad A\cup\overline{A}=\Omega.$$

Ясно также, что

$$A\cup\varnothing=A, \quad A\varnothing=\varnothing, \quad A\cup\Omega=\Omega, \quad A\Omega=A.$$

4.6. Свойства действий над событиями. Легко видеть, что для любых событий A и B событие $\overline{A}\cup\overline{B}$ противоположно событию AB:

$$\overline{A}\cup\overline{B}=\overline{AB}.$$

Действительно, $\overline{A}\cup\overline{B}$ есть появление хотя бы одного из событий $\overline{A}$ и $\overline{B}$, что равноценно непоявлению AB. И вообще для любого множества событий A_s, $s\in S$,

$$\bigcup_{s\in S}\overline{A}_s=\overline{\bigcap_{s\in S}A_s}.$$

Событие $\overline{A}\overline{B}$ представляет собой совместное появление $\overline{A}$ и $\overline{B}$, т. е. противоположно появлению хотя бы одного из событий A или B:

$$\overline{A}\overline{B}=\overline{A\cup B}.$$

И вообще для любого множества событий A_s, $s\in S$,

$$\bigcap_{s\in S}\overline{A}_s=\overline{\bigcup_{s\in S}A_s}.$$

Последние четыре формулы выражают п р и н ц и п д в о й с т в е н н о с т и: *операции объединения и пересечения меняются местами при переходе к противоположным событиям.*

Из установленных свойств операций пересечения и объединения следует, что для любых событий A и B

$$A=A\Omega=A(B\cup\overline{B})=AB\cup A\overline{B}. \tag{9}$$

Эта формула дает разложение любого события A на два непересекающихся события.

Если событие A обязательно происходит при появлении некоторого другого события B, то говорят, что событие B представляет собой *часть* или *подсобытие* события A и пишут $B\subset A$ или $A\supset B$.

Если события A и B могут появиться или не появиться только вместе, т. е. $B\subset A$ и $A\subset B$, то они на-

зываются *эквивалентными*, что обозначают равенством $A=B$.

Если $B\subset A$, то $AB=B$ и формула (9) принимает вид

$$A=B\cup A\overline{B}. \qquad (10)$$

4.7. Элементарные события. Читатель, знакомый с элементами теории множеств, легко заметит, что операции над событиями тождественны операциям над множествами. Эта аналогия между событиями и множествами объясняется тем, что каждое событие связано с определенным множеством исходов опыта так, что оно обязательно происходит при появлении одного из исходов, принадлежащих этому множеству, и не происходит при появлении одного из исходов, не принадлежащих этому множеству. Так, например, в схеме случаев п. 3.2 событие представляет собой объединение всех благоприятствующих ему случаев. В схеме геометрических вероятностей п. 3.3 исходом каждого опыта является попадание в определенную точку, а каждое событие представляет собой попадание на определенное множество точек.

Чтобы поставить эти факты на строгую математическую основу, вводят понятие элементарного события. *Элементарным событием* называется событие, не содержащее никаких подсобытий, кроме невозможного события и самого себя. Иными словами, элементарное событие представляет собой событие, которое не разделяется на несовместные события, ни одно из которых не является невозможным.

§ 5. Аксиомы теории вероятностей

5.1. Пространство элементарных событий. Мы пришли к понятию вероятности, отправляясь от понятия частоты события. Поэтому естественно считать, что вероятности должны обладать всеми свойствами частот, вытекающими из их определения. Для вероятностей эти свойства в общем случае не могут быть выведены *). Поэтому основные свойства вероятностей приходится принять как аксиомы.

На основании изложенного в §§ 2 и 4 мы приходим к следующей модели случайных явлений. С данным опы-

---

*) Для вероятностей эти свойства можно вывести только в частном случае конечного числа равновозможных несовместных исходов опыта, рассмотренном в п. 3.2.

том связано некоторое множество элементарных событий, так что в результате опыта обязательно появляется одно и только одно элементарное событие. Кроме того, с данным опытом связано некоторое множество событий, для которых определены вероятности.

Множество всех элементарных событий, связанных с данным опытом, называется *пространством элементарных событий* и обычно обозначается Ω. При этом любое элементарное событие — точка пространства Ω — обозначается буквой ω.

Каждое событие представляет собой некоторое множество элементарных событий. В частности, любое элементарное событие ω представляет собой множество, состоящее из одного элемента (одноточечное множество). Достоверное событие представляет собой множество всех элементарных событий Ω. Невозможное событие представляет собой пустое множество $\varnothing$.

5.2. Поле событий. Множество событий, для которых определены вероятности, будем называть *полем событий* и обозначать $\mathcal{S}$.

Чтобы принятая модель случайных явлений могла служить для построения теории вероятностей, поле событий $\mathcal{S}$ должно обладать определенными свойствами.

Во-первых, если вероятность определена для некоторого события A, то естественно потребовать, чтобы она была определена и для противоположного события $\overline{A}$. Следовательно, множество $\mathcal{S}$ должно содержать вместе с любым входящим в него событием A и противоположное событие $\overline{A}$, т. е. если $A \in \mathcal{S}$, то и $\overline{A} \in \mathcal{S}$.

Во-вторых, если вероятность определена для некоторых событий A и B, то естественно потребовать, чтобы она была определена и для их пересечения AB, чтобы можно было определить условные вероятности. Поэтому множество $\mathcal{S}$ должно содержать наряду с любыми двумя входящими в него событиями A и B и их пересечение AB, т. е. если $A, B \in \mathcal{S}$, то и $AB \in \mathcal{S}$.

Множество событий, обладающее перечисленными двумя свойствами, называется *алгеброй событий*. Таким образом, поле событий $\mathcal{S}$ должно быть алгеброй событий.

Изучим основные свойства поля событий, вытекающие из его определения как алгебры событий.

Пусть A и B — любые два события, принадлежащие полю $\mathcal{S}$. По определению противоположные события $\overline{A}$

и $\overline{B}$ и их пересечение $\overline{A}\overline{B}$ также принадлежат полю $\mathscr{S}$. Но тогда и событие $\overline{\overline{A}\overline{B}}$, противоположное событию $\overline{A}\overline{B}$, принадлежит полю $\mathscr{S}$. Но событие, противоположное пересечению $\overline{A}\overline{B}$, согласно принципу двойственности совпадает с объединением событий, противоположных событиям $\overline{A}$ и $\overline{B}$, т. е. с событием $A \cup B$, $\overline{\overline{A}\overline{B}} = A \cup B$. Следовательно, поле $\mathscr{S}$ содержит наряду с любыми двумя входящими в него событиями A и B и их объединение $A \cup B$.

Из этого свойства следует, что поле событий $\mathscr{S}$ содержит достоверное событие Ω. Действительно, для любого события $A \in \mathscr{S}$ и $\overline{A} \in \mathscr{S}$. Следовательно, и $A \cup \overline{A} = \Omega \in \mathscr{S}$.

В силу ассоциативности операций пересечения и объединения событий поле событий $\mathscr{S}$ содержит любые конечные пересечения и объединения входящих в него событий.

Наконец, поле $\mathscr{S}$ содержит невозможное событие $\varnothing$ как противоположное достоверному событию Ω.

Рассмотренные свойства поля достаточны для большей части применений теории вероятностей. Однако для решения некоторых сложных задач, выдвигаемых практикой, необходимо потребовать, чтобы поле событий $\mathscr{S}$ обладало еще одним свойством, а именно оно должно содержать не только конечные, но и все счетные объединения входящих в него событий, т. е.

$$\bigcup_{k=1}^{\infty} A_k \in \mathscr{S}, \text{ если } A_k \in \mathscr{S} \qquad (k = 1, 2, \ldots).$$

Алгебра событий, обладающая таким свойством, называется *σ-алгеброй* или *борелевским полем* событий.

Очевидно, что σ-алгебра содержит также все счетные пересечения входящих в нее событий. Это непосредственно вытекает из принципа двойственности для случая счетного множества событий.

В простейших задачах п. 3.2 все элементарные события принадлежат полю событий $\mathscr{S}$. В общем случае поле событий может содержать, а может и не содержать элементарные события.

5.3. Аксиомы. Перейдем к основным аксиомам теории вероятностей.

Аксиома 1. *Каждому событию $A \in \mathscr{S}$ соответствует неотрицательное число — вероятность этого события $P(A)$.*

Аксиома 2. *Вероятность достоверного события равна* 1: $P(\Omega)=1$.

Аксиома 3 (аксиома сложения вероятностей). *Вероятность объединения несовместных событий равна сумме их вероятностей*:

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i), \text{ если } A_k A_h=\varnothing \text{ при } k \neq h. \qquad (11)$$

Это справедливо как для конечного числа событий $A_1, \ldots, A_n$, так и для счетного множества событий $\{A_k\}$ (при $n=\infty$) *).

Для простейших задач теории вероятностей, к которым относятся, в частности, все задачи с конечным множеством равновозможных исходов опыта, рассмотренные в п. 3.2, достаточно принять аксиому сложения вероятностей только для любого конечного множества событий $A_1, \ldots, A_n$. Однако уже для задач с геометрическими вероятностями, рассмотренных в п. 3.3, несмотря на их простоту, необходимо принять аксиому сложения вероятностей для любых счетных множеств событий $\{A_k\}$.

Построение теории вероятностей на основе сформулированных трех аксиом принадлежит А. Н. Колмогорову, работы которого положили начало созданию современной теории вероятностей как строгой математической науки [39, 40].

5.4. Вероятность как функция множества — мера. Поскольку каждое событие $A \in \mathcal{S}$ представляет собой множество элементарных событий, вероятность $P(A)$ представляет собой функцию множества A, определенную на поле множеств $\mathcal{S}$.

Функции множества, обладающие свойством (11) при всех конечных n, называются *аддитивными*. Функции множества, обладающие свойством (11) при $n=\infty$, называются *счетно-аддитивными* или, короче, *σ-аддитивными* или *мерами*. Любая σ-аддитивная функция множества является аддитивной, однако не всякая аддитивная функция множества σ-аддитивна.

---

*) Ясно, что свойство сложения частот справедливо только для конечного числа событий $A_1, \ldots, A_n$, так как при любом конечном числе опытов может появиться только конечное число несовместных событий. Распространение соответствующего свойства вероятностей на счетное множество событий необходимо для построения достаточно общей теории, охватывающей все задачи, встречающиеся на практике.

В силу аксиом 1 и 3 вероятность $P(A)$ представляет собой неотрицательную σ-аддитивную функцию множества, т. е. неотрицательную меру.

5.5. Вероятностное пространство. Пространство элементарных событий Ω с заданной в нем алгеброй или σ-алгеброй множеств $\mathcal{S}$ и определенной на $\mathcal{S}$ вероятностью — неотрицательной мерой $P(A)$, $A \in \mathcal{S}$, называется *вероятностным пространством* и обозначается $(\Omega, \mathcal{S}, P)$. Таким образом, математической моделью любого случайного явления в современной теории вероятностей служит вероятностное пространство.

Соответствие между событиями некоторого множества событий и их вероятностями обычно называют *распределением вероятностей*. Таким образом, вероятность $P(A)$ как функция множества $A \in \mathcal{S}$ определяет распределение вероятностей на $\mathcal{S}$.

Пример 9. В любой задаче определения геометрических вероятностей пространством элементарных событий Ω служит множество всех точек соответствующего пространства (в частности, прямой или плоскости), в которые возможно попадание. Так в примере 3 Ω есть интервал $(0, 60c)$, в примере 4 Ω есть проекция самолета на плоскость, перпендикулярную траектории снаряда, в примере 5 Ω есть прямоугольник, образованный интервалом $[0, a]$ на оси x и интервалом $[0, \pi/2]$ на оси θ. Полем $\mathcal{S}$ в этом случае служит множество всех подмножеств A множества Ω, имеющих меру (длину, площадь, объем). Вероятность $P(A)$ определяется формулой

$$P(A) = v(A)/v(\Omega),$$

где $v(A)$ и $v(\Omega)$ — меры (длины, площади, объемы) множеств (частей пространства) A и Ω соответственно.

Распределение вероятностей можно определить вероятностями всех элементарных событий только в случае конечного или счетного множества элементарных событий. В большей же части задач, в частности в простейших задачах определения вероятностей попадания в различные части того или иного пространства, множество элементарных событий несчетно, причем, как правило, ни одно из них не может считаться существенно «более вероятным», чем любое другое. Вследствие этого вероятности всех элементарных событий в таких случаях равны нулю и потому ни в какой мере не определяют распределение вероятностей. Именно поэтому распределение вероятностей в общем случае определяется функцией множества $P(A)$.

Рассмотренная модель охватывает все задачи современной теории вероятностей. В каждой вероятностной

задаче можно определить некоторое вероятностное пространство. Правда, пространство элементарных событий часто оказывается очень сложным. Однако для развития теории вероятностей и для ее приложений конкретная структура пространства элементарных событий не существенна, вследствие чего можно не задавать это пространство в явном виде, а ограничиться лишь предположением о его существовании.

Понятие вероятностного пространства дает возможность положить в основу построения теории вероятностей методы теории множеств, теории меры и функционального анализа. В частности, все выводимые дальше свойства вероятностей и многие другие, которыми приходится пользоваться для построения более сложных разделов теории вероятностей — теории случайных функций и др., непосредственно вытекают из общих свойств меры.

5.6. Свойства вероятностей. Изучим теперь свойства вероятностей, вытекающие из аксиом.

Так как невозможное событие $\varnothing$ несовместно с любым другим событием A, $A\varnothing = \varnothing$, то из (11) следует, что $P(A \cup \varnothing) = P(A) + P(\varnothing)$. С другой стороны, так как $A \cup \varnothing = A$ (добавление невозможного события не изменяет события A), то $P(A \cup \varnothing) = P(A)$. Следовательно, $P(\varnothing) = 0$, т. е. *вероятность невозможного события равна нулю.*

Если $B \subset A$, то, представив A разложением (10) на два несовместных события, $A = B + A\bar{B}$, получим в силу (11) $P(A) = P(B) + P(A\bar{B})$, откуда следует, что $P(B) \leqslant P(A)$.

Таким образом, *если событие B может произойти только вместе с событием A, то вероятность события B не может быть больше вероятности события A.*

А так как любое событие A может произойти только вместе с достоверным событием Ω, $A = A\Omega \subset \Omega$, то никакое событие не может иметь вероятность, большую вероятности достоверного события, т. е. 1. Таким образом, *вероятность любого события принадлежит интервалу* $[0, 1]$:

$$0 \leqslant P(A) \leqslant 1.$$

Представив объединение совместных событий $A_1, \ldots, A_n$ в виде объединения несовместных событий:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2\bar{A}_1 \cup A_3\bar{A}_1\bar{A}_2 \cup \ldots \cup A_n\bar{A}_1 \ldots \bar{A}_{n-1},$$

получим на основании (11)

$$P\left(\bigcup_{i=1}^{n} A_i\right)=P(A_1)+P(A_2\overline{A_1})+P(A_3\overline{A_1}\overline{A_2})+\ldots$$
$$\ldots+P(A_n\overline{A_1}\ldots\overline{A_{n-1}}). \qquad (12)$$

А так как

$$A_2\overline{A_1}\subset A_2,\ A_3\overline{A_1}\overline{A_2}\subset A_3,\ \ldots,\ A_n\overline{A_1}\ldots\overline{A_{n-1}}\subset A_n,$$

то

$$P(A_2\overline{A_1})\leqslant P(A_2),\ \ldots,\ P(A_n\overline{A_1}\ldots\overline{A_{n-1}})\leqslant P(A_n),$$

и мы получаем

$$P\left(\bigcup_{i=1}^{n} A_i\right)\leqslant\sum_{i=1}^{n}P(A_i). \qquad (13)$$

Это неравенство, называемое *свойством полуаддитивности* вероятности, справедливо для любого конечного или счетного множества событий $\{A_k\}$ (т. е. как при конечном n, так и при $n=\infty$).

Представляет интерес вычислить вероятность объединения совместных событий. Применяя формулу (12) для двух событий $A_1=A$ и $A_2=B$, получим

$$P(A\cup B)=P(A)+P(B\overline{A}).$$

С другой стороны, на основании (9) $B=BA\cup B\overline{A}$, и, следовательно,

$$P(B)=P(AB)+P(B\overline{A}).$$

Определив отсюда $P(B\overline{A})$ и подставив в предыдущее равенство, получим

$$P(A\cup B)=P(A)+P(B)-P(AB).$$

Таким образом, мы доказали теорему сложения вероятностей: *вероятность объединения любых двух событий равна сумме их вероятностей минус вероятность их пересечения.*

Предоставляем читателю самостоятельно вывести соответствующую формулу для вероятности объединения любого числа совместных событий (применив, например, метод индукции).

5.7. Полная группа событий. Совокупность событий $\{A_k\}$ (конечная или счетная) называется *полной группой* событий, если хотя бы одно из них обязательно появляется в результате опыта. Иными словами, события

$A_1, \ldots, A_n$, $n \leqslant \infty$, образуют полную группу, если их объединение есть достоверное событие: $\bigcup A_k = \Omega$.

Из аксиомы сложения вероятностей следует, что *если события $A_1, \ldots, A_n$ несовместны и образуют полную группу, то сумма их вероятностей равна единице*:

$$\sum_{i=1}^{n} P(A_i) = 1. \tag{14}$$

Противоположные события несовместны и образуют полную группу. Поэтому из (14) следует, что *сумма вероятностей противоположных событий равна единице*:

$$P(A) + P(\bar{A}) = 1. \tag{15}$$

Эта формула очень важна для практики. Во многих задачах вероятность интересующего нас события трудно вычислить, в то время как вероятность противоположного события вычисляется очень легко. В таких случаях формула (15) дает вероятность интересующего нас события.

§ 6. Условные вероятности

6.1. Условная вероятность. Свойство умножения частот дает естественный способ определения условной вероятности.

Условной вероятностью $P(A \mid B)$ события A относительно события B в случае, когда $P(B) \neq 0$, называется отношение вероятности пересечения событий A и B к вероятности события B:

$$P(A \mid B) = \frac{P(AB)}{P(B)}. \tag{16}$$

При таком определении условной вероятности теорема умножения частот, очевидно, распространяется и на вероятности:

$$P(AB) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B). \tag{17}$$

Таким образом, *вероятность совместного появления двух событий равна вероятности одного из них, умноженной на условную вероятность другого*.

Из определения (16) следует, что условные вероятности различных событий относительно одного и того же события B, $P(B) \neq 0$, удовлетворяют аксиомам 1, 2 и 3.

Следовательно, вся развиваемая дальше теория справедлива и для условных вероятностей.

Из (17) следует, что *вероятность совместного появления любого числа событий равна вероятности одного из них, умноженной на условную вероятность другого относительно первого, на условную вероятность третьего относительно пересечения двух первых и т. д., на условную вероятность последнего относительно пересечения всех предыдущих*:

$$P(A_1 A_2 \ldots A_n) = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 A_2) \ldots \\ \ldots P(A_n \mid A_1 A_2 \ldots A_{n-1}). \quad (18)$$

Это следствие легко выводится из (17) по индукции.

Пример 10. В урне 12 шаров — 5 белых и 7 черных. Из урны вынимают два шара. Найти вероятность того, что оба шара окажутся белыми.

Введем события: A — первый (или условно считающийся первым, если шары вынимаются одновременно) шар белый, B — второй шар белый. Тогда будем иметь

$$P(A) = P(B) = 5/12, \quad P(B \mid A) = P(A \mid B) = 4/11$$

и по формуле (17)

$$P(AB) = \frac{5}{12} \cdot \frac{4}{11} = \frac{5}{33}.$$

Пример 11. В урне 16 шаров — 5 белых, 7 черных и 4 красных. Найти вероятность того, что среди вынутых из урны четырех шаров первый будет белым, второй — черным, а остальные два — красными.

Введем события: A_1 — первый шар белый, A_2 — второй шар черный, A_3 — третий шар красный, A_4 — четвертый шар красный. Тогда получим

$$P(A_1) = 5/16, \quad P(A_2 \mid A_1) = 7/15, \quad P(A_3 \mid A_1 A_2) = 4/14, \\ P(A_4 \mid A_1 A_2 A_3) = 3/13$$

и по формуле (10)

$$P(A_1 A_2 A_3 A_4) = \frac{5}{16} \cdot \frac{7}{15} \cdot \frac{4}{14} \cdot \frac{3}{13} = \frac{1}{104}.$$

Чтобы убедиться, что эта вероятность не зависит от того, в каком порядке берутся события, возьмем эти события в другом порядке, например A_3, A_1, A_4, A_2. Тогда получим

$$P(A_3) = 4/16, \quad P(A_1 \mid A_3) = 5/15, \\ P(A_4 \mid A_3 A_1) = 3/14, \quad P(A_2 \mid A_3 A_1 A_4) = 7/13, \\ P(A_3 A_1 A_4 A_2) = \frac{4}{16} \cdot \frac{5}{15} \cdot \frac{3}{14} \cdot \frac{7}{13} = \frac{1}{104}.$$

Пример 12. В теории надежности *функцией надежности* или просто *надежностью* $p(t)$ обычно называют вероятность исправной

работы прибора от момента $t=0$ до текущего момента t. В качестве исходной характеристики прибора при этом принимают *интенсивность отказов*

$$\lambda=\lim_{\Delta t\to 0}\frac{p(t+\Delta t\mid t)}{\Delta t},$$

представляющую собой предел отношения условной вероятности отказа прибора, исправно работавшего до момента t в интервале времени $(t,\ t+\Delta t)$, к величине этого интервала Δt при $\Delta t\longrightarrow 0$. Интенсивность отказов λ определяется экспериментально, по крайней мере для простейших элементов и блоков. Задача состоит в том, чтобы, зная интенсивность отказов как функцию времени $\lambda=\lambda(t)$, найти функцию надежности $p(t)$.

Для решения задачи назовем отказ прибора в интервале времени $(t,\ t+\Delta t)$ событием A, а его исправную работу до момента t — событием B. Исправную работу прибора до момента $t+\Delta t$ назовем событием C. Событие C можно выразить через события A и B. Для того чтобы система исправно работала до момента $t+\Delta t$, необходимо, чтобы она работала исправно до момента t и чтобы она работала исправно в интервале времени от t до $t+\Delta t$. Следовательно, событие C представляет собой пересечение двух событий: события B и события, противоположного событию A, т. е. $C=B\overline{A}$. Отсюда на основании (17) следует, что

$$P(C)=P(B)\,P(\overline{A}\mid B). \tag{I}$$

Но $P(B)$ как вероятность исправной работы системы до момента t представляет собой искомую функцию надежности $p(t)$:

$$P(B)=p(t). \tag{II}$$

Вероятность $P(C)$, т. е. вероятность исправной работы системы до момента $t+\Delta t$, представляет собой значение той же функции надежности в момент $t+\Delta t$:

$$P(C)=p(t+\Delta t). \tag{III}$$

Наконец, $P(A\mid B)$ представляет собой условную вероятность отказа системы в интервале времени $(t,\ t+\Delta t)$, которая выражается через интенсивность отказов системы λ формулой

$$P(A\mid B)=P(t+\Delta t\mid t)=\lambda\,\Delta t+o(\Delta t),$$

где $o(\Delta t)$, как всегда, означает бесконечно малую высшего порядка по сравнению с Δt. Применяя для вычисления условной вероятности противоположного события $\overline{A}$ формулу (15), найдем

$$P(\overline{A}\mid B)=1-\lambda\,\Delta t+o(\Delta t). \tag{IV}$$

Подставляя выражения (II), (III) и (IV) в (I), получим

$$p(t+\Delta t)=p(t)(1-\lambda\,\Delta t)+o(\Delta t).$$

Теперь остается раскрыть в правой части полученного равенства скобки, перенести $p(t)$ налево, разделить все члены на Δt и перейти к пределу при $\Delta t\longrightarrow 0$. При этом слева получим производную $p'(t)$. В результате получается дифференциальное уравнение для функции надежности

$$p'(t)=-\lambda p(t). \tag{V}$$

В качестве начального условия естественно принять условие, что система начинает работать в исправном состоянии, т. е. $p(0)=1$. Дифференциальное уравнение (V) и начальное условие $p(0)=1$ полностью определяют искомую функцию надежности $p(t)$.

Легко проверить непосредственной подстановкой, что интеграл уравнения (V), равный единице в начальный момент, определяется формулой

$$p(t)=\exp\left\{-\int_0^t \lambda\, d\tau\right\}. \tag{VI}$$

6.2. Зависимые и независимые события. Два события называются *независимыми*, если появление одного из них не изменяет вероятности другого, или, иными словами, если появление одного из них не содержит никакой информации о другом.

События A и B называются *зависимыми*, если появление одного из них изменяет вероятность другого.

Для независимых событий A и B

$$P(A\mid B)=P(A), \qquad P(B\mid A)=P(B). \tag{19}$$

Действительно, вероятность события A, когда известно, что B произошло, есть условная вероятность A относительно B. Если при появлении события B вероятность события A не изменяется, то это означает, что условная вероятность события A относительно B совпадает с вероятностью события A, которую, в отличие от условной вероятности, называют *безусловной*.

Очевидно, что для независимости событий A и B достаточно одного из двух равенств (19); второе автоматически будет выполнено в силу (17).

Для зависимых событий A и B

$$P(A\mid B)\neq P(A), \qquad P(B\mid A)\neq P(B). \tag{20}$$

Очевидно, что два несовместных события A и B всегда зависимы, так как появление одного из них исключает другое, вследствие чего $P(A\mid B)=P(B\mid A)=0$.

События $A_1, \ldots, A_n$ называются *независимыми*, если каждое из них не зависит от каждого из остальных и от всех возможных их пересечений.

Заметим, что для независимости событий $A_1, \ldots, A_n$ их попарная независимость необходима, но не достаточна.

Пример 13. Пусть A и B — произвольные независимые события, вероятности которых равны $1/2$, и $C=AB\cup\bar{A}\bar{B}$. Очевидно, что

$$P(C)=P(AB)+P(\bar{A}\bar{B})=P(A)P(B)+P(\bar{A})P(\bar{B})=1/2,$$
$$P(C\mid A)=P(B)=1/2, \quad P(C\mid B)=P(A)=1/2.$$

Таким образом, A, B и C попарно независимы. Однако $P(C \mid AB)=1$, так как если происходит событие AB, то происходит и C. Следовательно, события A, B и C зависимы. Предлагаем читателю самостоятельно вычислить остальные условные вероятности в этой задаче.

6.3. Теорема умножения вероятностей для независимых событий. Рассмотрим случай независимых событий $A_1, \ldots, A_n$.

▶ Если события $A_1, \ldots, A_n$ независимы, то события A_n и $A_1A_2\ldots A_{n-1}$ независимы, события A_{n-1} и $A_1A_2\ldots A_{n-2}$ независимы, и так далее, события A_3 и A_1A_2 независимы, и, наконец, события A_1 и A_2 независимы. Поэтому

$$P(A_n \mid A_1A_2\ldots A_{n-1}) = P(A_n),$$
$$P(A_{n-1} \mid A_1A_2\ldots A_{n-2}) = P(A_{n-1}),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$P(A_3 \mid A_1A_2) = P(A_3),$$
$$P(A_2 \mid A_1) = P(A_2),$$

и формула (18) принимает вид

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2)\ldots P(A_n). \blacktriangleleft \qquad (21)$$

Таким образом, *вероятность пересечения независимых событий равна произведению их вероятностей.*

Пример 14. Вероятность попадания в цель при сбрасывании одной бомбы равна $p=0{,}1$. Найти вероятность попадания в цель хотя бы одной бомбы, если бомбометание по цели производят 10 самолетов, каждый из которых сбрасывает одну бомбу.

Пусть событие B представляет собой попадание хотя бы одной бомбы в цель. Противоположное событие $\overline{B}$ представляет собой непопадание ни одной бомбы. В данном примере попадания в цель разных бомб можно считать независимыми событиями, так как каждый самолет осуществляет бомбометание независимо от других. Следовательно, на основании формулы (21) вероятность того, что ни одна бомба не попадет в цель, равна

$$P(\overline{B}) = (1-p)^{10}.$$

Для определения искомой вероятности $P(B)$ попадания хотя бы одной бомбы в цель теперь достаточно воспользоваться формулой (15), связывающей вероятности противоположных событий:

$$P(B) = 1-(1-p)^{10} = 1-0{,}9^{10} \approx 0{,}63.$$

Непосредственное вычисление этой вероятности, как мы увидим в п. 8.4, очень сложно.

Пример 15. В урне 12 шаров — 5 белых и 7 черных. Из урны вынимают один шар, отмечают его цвет и возвращают в урну. После этого шары тщательно перемешивают и из урны вынимают второй шар. Найти вероятность того, что оба раза появится белый шар.

В данном случае, поскольку после первого вынимания шар возвращают в урну, информация о появлении белого шара при первом

вынимании не изменяет вероятности появления белого шара при следующем. Поэтому события A (появление белого шара в первый раз) и B (появление белого шара во второй раз) независимы и вероятность их совместного появления равна произведению вероятностей:

$$P(AB)=P(A)P(B)=\frac{5}{12}\cdot\frac{5}{12}=\frac{25}{144}.$$

Пример 16. В урне 16 шаров — 5 белых, 7 черных и 4 красных. Из урны 4 раза вынимают по одному шару, возвращая каждый раз шар в урну. Найти вероятность того, что первый шар будет белым, второй — черным, а третий и четвертый — красными.

В данном случае события A_1 — первый шар белый, A_2 — второй шар черный, A_3 — третий шар красный, A_4 — четвертый шар красный независимы. Поэтому

$$P(A_1A_2A_3A_4)=\frac{5}{12}\cdot\frac{7}{12}\cdot\frac{5}{12}\cdot\frac{4}{12}=\frac{35}{1296}.$$

§ 7. Вероятности сложных событий

7.1. Формула полной вероятности. Предположим, что с данным опытом связана полная группа несовместных событий $H_1, \ldots, H_n$, вероятности которых $P(H_i)$ $(i=1, \ldots, n)$ известны. Нас интересует событие A, для которого известны условные вероятности $P(A|H_i)$ $(i=1, \ldots, n)$ относительно всех событий $H_1, \ldots, H_n$. Требуется найти вероятность события A.

► Эта задача решается очень просто. Поскольку события $H_1, \ldots, H_n$ образуют полную группу, их объединение есть достоверное событие. Событие A может появиться только вместе с каким-нибудь событием H_k. Таким образом, событие A есть объединение событий $AH_1, \ldots, AH_n$. Так как события $H_1, \ldots, H_n$ по условию несовместны, то события $AH_1, \ldots, AH_n$ тоже несовместны, и мы можем применить аксиому сложения:

$$P(A)=P\left(\bigcup_{i=1}^{n} AH_i\right)=\sum_{i=1}^{n} P(AH_i).$$

Отсюда, пользуясь формулой (17), получаем

$$P(A)=\sum_{i=1}^{n} P(H_i)P(A|H_i). \quad ◄ \qquad (22)$$

Итак, *вероятность события A равна сумме вероятностей событий $H_1, \ldots, H_n$, умноженных на соответствующие условные вероятности события A.*

Формула (22) называется *формулой полной вероятности*. Она широко применяется в теории вероятностей и в ее приложениях.

Пример 17. На эксплуатацию приходят однотипные приборы, выпускаемые тремя заводами в пропорции $n_1:n_2:n_3$ (т. е. на каждые n_1 приборов, выпускаемых первым заводом, приходятся n_2 приборов, выпускаемых вторым заводом, и n_3 приборов, выпускаемых третьим заводом). Предположим, что марка завода на приборах отсутствует и что интенсивностъ отказов (постоянная) для приборов первого завода равна λ_1, для приборов второго завода — λ_2 и для приборов третьего завода — λ_3. Найти функцию надежности данного прибора.

В данной задаче имеем три несовместных события, образующих полную группу:

H_1 — данный прибор выпущен первым заводом,
H_2 — данный прибор выпущен вторым заводом,
H_3 — данный прибор выпущен третьим заводом.

Вероятности этих событий легко определяются:

$$P(H_k) = \frac{n_k}{n_1 + n_2 + n_3} \qquad (k = 1, 2, 3).$$

Условные вероятности интересующего нас события A — исправной работы прибора до момента t — относительно событий H_1, H_2, H_3 легко определяются по формуле (VI) примера 11:

$$P(A \mid H_1) = e^{-\lambda_1 t}, \quad P(A \mid H_2) = e^{-\lambda_2 t}, \quad P(A \mid H_3) = e^{-\lambda_3 t}.$$

Пользуясь формулой полной вероятности (22), находим вероятность исправной работы прибора до момента t (т. е. искомую функцию надежности прибора):

$$P(A) = p(t) = \frac{n_1 e^{-\lambda_1 t} + n_2 e^{-\lambda_2 t} + n_3 e^{-\lambda_3 t}}{n_1 + n_2 + n_3}.$$

Пример 18. Цель (например, самолет) может быть поражена огнем авиационных пушек при попадании в нее одного, двух, трех и т. д. снарядов. Найти вероятность поражения цели очередью из n снарядов, если известны вероятности попадания в нее одного, двух и т. д., n снарядов $P_{m,n}$ $(m = 1, \ldots, n)$ и условные вероятности поражения цели при попадании в нее одного, двух и т. д. n снарядов $G(m)$ $(m = 1, \ldots, n)$.

В данном случае имеем $n+1$ несовместных событий, образующих полную группу:

H_1 — попадание в цель одного снаряда,
H_2 — попадание в цель двух снарядов,
. . . .
H_n — попадание в цель n снарядов,
H_{n+1} — непопадание в цель ни одного снаряда.

Вероятности событий $H_1, \ldots, H_n$ по условию известны:

$$P(H_m) = P_{m,n} \qquad (m = 1, \ldots, n).$$

Вероятность события H_{n+1} может быть найдена по формуле (14), связывающей вероятности несовместных событий, образующих полную группу:

$$P(H_{n+1}) = 1 - P_{1,n} - P_{2,n} - \ldots - P_{n,n}.$$

По условию известны также условные вероятности интересующего нас события A — поражение цели — относительно событий $H_1, \ldots, H_n$:

$$P(A \mid H_m) = G(m) \qquad (m = 1, \ldots, n).$$

Условная вероятность события A относительно H_{n+1}, очевидно, равна нулю, так как цель не может быть поражена при отсутствии попадания в нее: $P(A \mid H_{n+1}) = 0$. Следовательно, вероятность поражения цели может быть вычислена по формуле полной вероятности (22):

$$P(A) = \sum_{m=1}^{n+1} P(H_m) P(A \mid H_m) = \sum_{m=1}^{n} P_{m,n} G(m).$$

7.2. Формула Бейеса. В задачах практики нас часто интересует полная группа несовместных событий $H_1, \ldots$ $\ldots, H_n$, вероятности которых $P(H_i)$ $(i = 1, \ldots, n)$ известны. Эти события непосредственно не наблюдаемы, но можно наблюдать некоторое событие A, с ними связанное, для которого известны условные вероятности $P(A \mid H_i)$ $(i = 1, \ldots, n)$. Допустим, что произведен опыт, в результате которого появилось событие A. На основании этого опыта требуется сделать выводы относительно событий $H_1, \ldots$ $\ldots, H_n$, т. е. определить, как изменились их вероятности после произведенного опыта. Иначе говоря, нужно найти условные вероятности событий $H_1, \ldots, H_n$ относительно события A.

► На основании теоремы умножения вероятностей (17)

$$P(AH_k) = P(A) P(H_k \mid A) = P(H_k) P(A \mid H_k).$$

Отсюда следует

$$P(H_k \mid A) = \frac{P(H_k) P(A \mid H_k)}{P(A)}.$$

Подставляя сюда выражение вероятности события A из формулы полной вероятности (22), получим

$$P(H_k \mid A) = \frac{P(H_k) P(A \mid H_k)}{\sum_{i=1}^{n} P(H_i) P(A \mid H_i)} \qquad (k = 1, \ldots, n). \blacktriangleleft \quad (23)$$

Эта формула обычно называется *формулой Бейеса*. Она решает поставленную задачу.

Вероятности $P(H_k)$ $(k = 1, \ldots, n)$ интересующих нас событий $H_1, \ldots, H_n$ до опыта обычно называются *априорными* вероятностями от латинского *a priori*, что значит «сперва», т. е. в данном случае до того, как был произведен опыт. Вероятности $P(H_k \mid A)$ $(k = 1, \ldots, n)$ тех же

событий после опыта называются *апостериорными* от латинского *a posteriori*, что значит «после», т. е. в данном случае после опыта.

Пример 19. В условиях примера 17 предположим, что прибор проработал безотказно время T (событие A). Найти апостериорную вероятность того, что прибор выпущен k-м заводом ($k=1, 2, 3$).

Подставив вероятности $P(H_1)$, $P(H_2)$, $P(H_3)$, $P(A \mid H_1)$, $P(A \mid H_2)$, $p(A \mid H_3)$, заданные в примере 17, в формулу Бейеса (23), находим апостериорную вероятность того, что прибор выпущен k-м заводом:

$$P(H_k \mid A) = \frac{n_k e^{-\lambda_k T}}{n_1 e^{-\lambda_1 T} + n_2 e^{-\lambda_2 T} + n_3 e^{-\lambda_3 T}}.$$

§ 8. Повторение опытов

8.1. Случай постоянных условий опыта. Рассмотрим сложный опыт, состоящий из нескольких более простых опытов, в каждом из которых может появиться или не появиться некоторое событие A.

Опыты называются *независимыми*, если вероятность интересующего нас события A в каждом опыте не зависит от результатов других опытов. Предположим, что производятся n независимых опытов, в каждом из которых вероятность события A равна p. Требуется найти вероятность $P_{m,n}$ того, что событие A появится m раз.

▶ Для того чтобы при n опытах событие A появилось m раз, необходимо и достаточно, чтобы появилась одна из последовательностей событий $B_1, \ldots, B_n$, в которых m из событий $B_1, \ldots, B_n$ совпадают с A, а $n-m$ — с противоположным событием $\bar{A}$. Очевидно, что число таких последовательностей равно числу сочетаний из n по m, т. е. $C_n^m = n!/[m!(n-m)!]$, $0! = 1$. В силу независимости опытов вероятность каждой такой последовательности по теореме умножения для независимых событий (21) равна $p^m q^{n-m}$, где $q = 1 - p$. Наконец, в силу несовместности всех возможных последовательностей искомая вероятность $P_{m,n}$ равна сумме вероятностей всех последовательностей, состоящих из m событий A и $n-m$ событий $\bar{A}$, т. е. сумме C_n^m слагаемых, равных $p^m q^{n-m}$:

$$P_{m,n} = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m} \quad (m = 0, 1, \ldots, n). \blacktriangleleft \tag{24}$$

Возьмем теперь вспомогательную переменную u и заметим, что величина $P_{m,n} u^m = C_n^m p^m q^{n-m} u^m$ представляет

собой общий член разложения функции $(q+pu)^n$ по формуле бинома Ньютона.

Таким образом, вероятность $P_{m,n}$ представляет собой коэффициент при u^m в разложении функции

$$\varphi_n(u) = (q+pu)^n \tag{25}$$

по степеням u.

Функция $\varphi_n(u)$ называется *производящей функцией* для вероятностей $P_{m,n}$.

К полученному результату можно прийти также методом индукции, заметив, что при $n=1$ задача имеет очевидное решение $P_{0,1}=q$, $P_{1,1}=p$, и применив известное соотношение между биномиальными коэффициентами

$$C_{n-1}^{m-1} + C_{n-1}^{m} = C_n^m. \tag{26}$$

Можно также, исходя из того, что

$$\varphi_1(u) = q + pu, \tag{27}$$

вывести из (11) и (21) рекуррентную формулу для производящих функций:

$$\varphi_n(u) = \varphi_{n-1}(u)(q+pu), \tag{28}$$

а затем получить (25).

Элементарными событиями в данном случае служат все конечные последовательности $\{B_1, \ldots, B_n\}$, где каждое B_k представляет собой событие A или противоположное событие $\bar{A}$. Полем событий $\mathcal{S}$ служит алгебра всех возможных объединений этих элементарных событий, дополненных невозможным событием. Вероятность каждого элементарного события равна $p^m q^{n-m}$, где m — число событий B_k в последовательности $\{B_1, \ldots, B_n\}$, совпадающих с A $(m=0, 1, \ldots, n)$. Вероятность любого события определяется как сумма вероятностей входящих в него элементарных событий.

Соответствие между числами $m=0, 1, \ldots, n$ и вероятностями $P_{m,n}$, определяемое формулой (24), называется *биномиальным распределением*.

8.2. Случай переменных условий опыта. Формулы (24) и (25) легко обобщаются на случай, когда вероятность события A имеет различные значения в разных опытах (повторение опытов в переменных условиях). Если опыты независимы и вероятность события A в k-м опыте равна p_k, $q_k = 1 - p_k$ $(k=1, \ldots, n)$, то вместо (24) совершенно

таким же путем получается формула

$$P_{m,n} = \sum p_{i_1} \dots p_{i_m} q_{i_{m+1}} \dots q_{i_n} \qquad (m = 0, 1, \dots, n), \quad (29)$$

где сумма распространена на все возможные разделения чисел $1, 2, \dots, n$ на две группы, одна из которых содержит m чисел $(i_1, i_2, \dots, i_m)$, а другая — $n-m$ чисел $(i_{m+1}, \dots, i_n)$. Число таких разделений равно C_n^m.

Легко понять, что вероятность $P_{m,n}$ в этом случае представляет собой коэффициент при u^m в разложении по степеням u производящей функции

$$\varphi_n(u) = \prod_{k=1}^{n} (q_k + p_k u). \quad (30)$$

К этому результату тоже можно прийти по индукции или получив рекуррентную формулу

$$\varphi_n(u) = \varphi_{n-1}(u)(q_n + p_n u). \quad (31)$$

8.3. Вероятность появления события не меньше данного числа раз. Во многих задачах практики приходится определять вероятность того, что интересующее нас событие A появится при n опытах не меньше чем данное число k раз.

► Очевидно, что сложное событие — появление события A не меньше чем k раз — представляет собой объединение $n-k+1$ несовместных событий: появление A ровно k раз, появление A ровно $k+1$ раз, и т. д., появление A ровно n раз. Следовательно, искомая вероятность $R_{k,n}$ того, что при n опытах событие A появится не меньше чем k раз, равна

$$R_{k,n} = P_{k,n} + P_{k+1,n} + \dots + P_{n,n} = \sum_{m=k}^{n} P_{m,n}. \quad (32)$$

Эту вероятность можно также вычислить, определив сначала вероятность противоположного события, т. е. вероятность того, что событие A появится меньше чем k раз, и вычтя ее из единицы:

$$R_{k,n} = 1 - P_{0,n} - P_{1,n} - \dots P_{k-1,n} = 1 - \sum_{m=0}^{k-1} P_{m,n}. \quad ◄ \quad (33)$$

Ясно, что всегда целесообразно пользоваться той из формул (32) и (33), у которой в сумме меньше слагаемых. Очевидно, что в (32) сумма содержит $n-k+1$ слагаемых, а в (33) — k слагаемых. Следовательно, форму-

лой (32) целесообразно пользоваться, если $n-k+1 \leqslant k$, т. е. если $k \geqslant (n+1)/2$. Если $k<(n+1)/2$, т. е. $k< < n-k+1$, то целесообразно пользоваться формулой (33).

8.4. Вероятность хотя бы одного появления события. Чаще всего приходится вычислять вероятность того, что интересующее нас событие появится хотя бы один раз (т. е. не меньше чем один раз). Очевидно, что в этом случае при любом $n \geqslant 2$ целесообразно пользоваться формулой (33), так как сумма в ней содержит только одно слагаемое $P_{0,n}$. В результате получим

$$R_{1,n}=1-P_{0,n}=1-q_1q_2\ldots q_n. \qquad (34)$$

В частном случае постоянных условий опыта $q_1=q_2=\ldots$ $\ldots=q_n=q$ и формула (34) принимает вид

$$R_{1,n}=1-q^n. \qquad (35)$$

Формулы (34) и (35) легко выводятся непосредственно без применения формулы (33). Для этого достаточно вычислить вероятность противоположного события — непоявления события A ни разу. Вероятность того, что при n независимых опытах событие A не появится ни разу, равна произведению вероятностей непоявления A в первом, втором, третьем, и т. д., n-м опытах, т. е. произведению $q_1q_2\ldots q_n$. Вычитая эту вероятность из единицы, мы получим формулу (34).

8.5. Случай опытов с несколькими событиями. Если в результате каждого опыта появляется одно из полной группы несовместных событий $A_1, \ldots, A_r$, вероятности которых равны соответственно $p_1, \ldots, p_r$, $p_1+\ldots+p_r=1$, то вероятность того, что при n опытах A_1 появится m_1 раз, A_2 — m_2 раз, и т. д., A_r — m_r раз, $m_1+\ldots+m_r=n$, определяется формулой

$$P_{m_1,\ldots,m_r}=\frac{n!}{m_1!\,m_2!\ldots m_r!}\,p_1^{m_1}p_2^{m_2}\ldots p_r^{m_r}. \qquad (36)$$

▶ Эта формула легко получается последовательным применением формул (29). Вероятность того, что событие A_1 появится m_1 раз при n опытах, согласно (29) равна

$$P_{m_1,n}=\frac{n!}{m_1!\,(m_2+\ldots m_r)!}\,p_1^{m_1}\,(p_2+\ldots+p_r)^{m_2+\ldots+m_r}.$$

Условная вероятность события A_2 в каждом из оставшихся опытов при условии, что A_1 в этих опытах не появляется, очевидно, равна $p_2/(p_2+\ldots+p_r)$. Поэтому

условная вероятность того, что в $m_2+\ldots+m_r$ опытах, в которых не появляется A_1, событие A_2 появится m_2 раз, согласно (29) определяется формулой

$$P_{m_2,\, m_2+\ldots+m_r}=\frac{(m_2+\ldots+m_r)!}{m_2!\,(m_3+\ldots+m_r)!}\times$$
$$\times\left(\frac{p_2}{p_2+\ldots+p_r}\right)^{m_2}\left(\frac{p_3+\ldots+p_r}{p_2+\ldots+p_r}\right)^{m_3+\ldots+m_r}.$$

Продолжая таким образом, находим условную вероятность того, что событие A_k появится m_k раз в $m_k+\ldots+m_r$ опытах при условии, что события $A_1, \ldots, A_{k-1}$ в этих опытах не появляются:

$$P_{m_k,\, m_k+\ldots+m_r}=\frac{(m_k+\ldots+m_r)!}{m_k!\,(m_{k+1}+\ldots+m_r)!}\times$$
$$\times\left(\frac{p_k}{p_k+\ldots+p_r}\right)^{m_k}\left(\frac{p_{k+1}+\ldots+p_r}{p_k+\ldots+p_r}\right)^{m_{k+1}+\ldots+m_r}.$$

Перемножая найденные вероятности, мы и получим формулу (36). ◀

Производящая функция вероятностей $P_{m_1,\ldots,m_r}$ $(m_1+\ldots+m_r=n)$ определяется формулой

$$\varphi_n(u_1, \ldots, u_r)=(p_1u_1+\ldots+p_ru_r)^n, \qquad (37)$$

причем $P_{m_1,\ldots,m_r}$ представляет собой коэффициент при $u_1^{m_1}\ldots u_r^{m_r}$ в разложении этой функции по степеням переменных $u_1, \ldots, u_r$.

Элементарными событиями в данном случае служат конечные последовательности $\{B_1, \ldots, B_n\}$, где каждое B_k представляет собой одно из событий $A_1, \ldots, A_n$. Полем событий служит алгебра всех возможных объединений этих событий, дополненных невозможным событием. Вероятность каждого элементарного события равна $p_1^{m_1}\ldots$ $\ldots p_r^{m_r}$, где m_i — число событий B_k в последовательности $\{B_1, \ldots, B_n\}$, совпадающих с A_i $(i=1, \ldots, r)$. Вероятность любого события равна сумме вероятностей входящих в него элементарных событий.

Распределение вероятностей, определяемое формулой (36), называется *полиномиальным*.

§ 9. Распределение Пуассона

9.1. Потоки событий. На практике приходится встречаться с событиями, которые происходят в случайные моменты времени. Такие события образуют последова-

тельность событий, называемую обычно *потоком событий*. Примерами потоков событий могут служить вызовы абонентов на телефонной станции, пересечения перекрестка транспортными средствами, вызовы скорой медицинской помощи, отказы технической системы, приходы клиентов в обслуживающую организацию (например, парикмахерскую) и т. д.

Часто можно считать, что поток событий удовлетворяет следующим условиям:

1) для любых двух непересекающихся интервалов времени вероятность появления любого данного числа событий в течение одного из них не зависит от того, сколько событий появляется в течение другого;

2) вероятность появления одного события в течение бесконечно малого интервала времени $(t, t+\Delta t)$ есть бесконечно малая величина порядка Δt;

3) вероятность появления более одного события в течение интервала времени $(t, t+\Delta t)$ есть бесконечно малая высшего порядка по сравнению с Δt.

Обозначим $p_m(t_1, t_2)$ вероятность появления m событий в интервале времени (t_1, t_2). Тогда условия 2) и 3) запишутся в виде

$$p_1(t, t+\Delta t) = \lambda(t)\Delta t + o(\Delta t), \tag{38}$$

$$\sum_{k=2}^{\infty} p_k(t+\Delta t) = o(\Delta t), \tag{39}$$

где $\lambda(t)$ — некоторая неотрицательная функция *).

9.2. Уравнение для вероятности непоявления событий. Поставим задачу: для потока событий, удовлетворяющего условиям 1), 2) и 3), найти вероятности того, что в данном интервале времени (t_0, t) появится m событий $(m = 0, 1, 2, \ldots)$.

► Считая момент t_0 фиксированным, обозначим искомые вероятности $p_m(t)$ $(m = 0, 1, 2, \ldots)$.

Для вычисления $p_0(t)$ заметим, что $p_0(t+\Delta t)$ представляет собой вероятность пересечения двух событий: ни одного события в интервале (t_0, t) и ни одного события в интервале $(t, t+\Delta t)$. Согласно условию 1) эти

---

*) $o(\Delta t)$, как всегда, означает бесконечно малую высшего порядка по сравнению с Δt, так что

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

события независимы. Поэтому

$$p_0(t+\Delta t) = p_0(t)\, p_0(t,\, t+\Delta t). \tag{40}$$

Но на основании (38) и (39)

$$p_0(t,\, t+\Delta t) = 1 - \sum_{k=1}^{\infty} p_k(t,\, t+\Delta t) = 1 - \lambda(t)\,\Delta t + o(\Delta t). \tag{41}$$

Подставив это выражение в (40), получим

$$p_0(t+\Delta t) = p_0(t) - p_0(t)\,\lambda(t)\,\Delta t + o(\Delta t),$$

откуда

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = -\lambda(t)\, p_0(t) + \frac{o(\Delta t)}{\Delta t}.$$

При $\Delta t \to 0$ правая часть этого равенства стремится к определенному пределу $-\lambda(t)\,p_0(t)$. Следовательно, существует и предел левой части. Таким образом, вероятность $p_0(t)$ дифференцируема при любом t и в пределе при $\Delta t \to 0$ мы получаем дифференциальное уравнение

$$p_0'(t) = -\lambda(t)\, p_0(t). \tag{42}$$

Для нахождения начального значения вероятности $p_0(t)$ достаточно положить в (41) $t = t_0$ и перейти к пределу при $\Delta t \to 0$. Тогда получим $p_0(t_0) = 1$.

9.3. Уравнения для вероятностей различных чисел событий. Для составления уравнений для вероятностей $p_1(t)$, $p_2(t)$, ... заметим, что m событий могут появиться в интервале времени $(t_0,\, t+\Delta t)$ одним из следующих $m+1$ несовместных способов: все m событий появляются в интервале $(t_0,\, t)$ и ни одного в интервале $(t,\, t+\Delta t)$, $m-1$ событий появляются в интервале $(t_0,\, t)$ и одно в интервале $(t,\, t+\Delta t)$, и т. д., все m событий появляются в интервале $(t,\, t+\Delta t)$. Поэтому на основании аксиомы сложения вероятностей и теоремы умножения вероятностей независимых событий (21) имеем

$$p_m(t+\Delta t) = p_m(t)\, p_0(t,\, t+\Delta t) + p_{m-1}(t)\, p_1(t,\, t+\Delta t) + \ldots$$
$$\ldots + p_0(t)\, p_m(t,\, t+\Delta t).$$

Отсюда на основании (38), (39) и (41) получаем

$$p_m(t+\Delta t) = p_m(t) + [p_{m-1}(t) - p_m(t)]\,\lambda(t)\,\Delta t + o(\Delta t).$$

Следовательно,

$$\frac{p_m(t+\Delta t) - p_m(t)}{\Delta t} = \lambda(t)\,[p_{m-1}(t) - p_m(t)] + \frac{o(\Delta t)}{\Delta t}.$$

Рассуждая далее совершенно так же, как и при выводе уравнения (42), получаем дифференциальное уравнение

$$p'_m(t) = \lambda(t)\,[p_{m-1}(t) - p_m(t)] \qquad (m = 1, 2, \ldots). \quad (43)$$

Начальные значения вероятностей $p_1(t)$, $p_2(t)$, ... все равны нулю в силу того, что $p_0(t_0) = 1$, $p_m(t_0) = 0$ ($m = 1$, 2, ...).

9.4. Решение уравнений. Приняв за независимую переменную величину

$$\mu = \int_{t_0}^{t} \lambda(\tau)\,d\tau, \quad (44)$$

приведем уравнения (42) и (43) к виду

$$\frac{dp_0}{d\mu} = -p_0, \quad \frac{dp_m}{d\mu} = -p_m + p_{m-1} \qquad (m = 1, 2, \ldots). \quad (45)$$

Начальные условия примут вид $p_0 = 1$, $p_m = 0$ ($m = 1$, 2, ...) при $\mu = 0$. Легко убедиться непосредственной подстановкой, что интегралы уравнений (45), удовлетворяющие начальным условиям, определяются формулой

$$p_m = \frac{\mu^m}{m!} e^{-\mu} \qquad (m = 0, 1, 2, \ldots). \blacktriangleleft \quad (46)$$

Таким образом, для данного интервала времени (t_0, t) мы имеем счетное множество элементарных событий: ни одного события в этом интервале, одно, два и т. д., и вероятности этих событий определяются формулой (46). Следовательно, формула (46) определяет распределение вероятностей. Это распределение вероятностей называется *распределением Пуассона*. Поэтому поток событий, удовлетворяющих условиям 1), 2) и 3), называется *пуассоновским потоком*. Параметр μ распределения Пуассона, как мы увидим в п. 3.2.1 (пример 3.2), представляет собой среднее число событий, происходящих в данном интервале времени (t_0, t). Функция $\lambda(t)$ называется *интенсивностью* пуассоновского потока.

Пример 20. Найти вероятность того, что число электронов, вылетающих из катода электронной лампы в течение интервала времени длительности t, будет равно m, если среднее число электронов, испускаемых в единицу времени, равно $\lambda = \text{const}$. Поток электронов можно считать пуассоновским.

На основании (44) в данном случае $\mu = \lambda t$. Подставив это выражение в (46), получим

$$p_m = \frac{(\lambda t)^m}{m!} e^{-\lambda t} \qquad (m = 0, 1, 2, \ldots).$$

Пример 21. Интенсивность потока вызовов телефонной станции (т. е. средняя плотность вызовов — предел отношения среднего числа вызовов в течение бесконечно малого интервала времени $(t, t+\Delta t)$ к Δt при $\Delta t \to 0$) равна $\lambda(t)$. Найти вероятность того, что в течение интервала времени (t_1, t_2) на станцию поступит m вызовов.

В данном случае можно с достаточной точностью считать, что поток вызовов удовлетворяет условиям, при которых поток является пуассоновским. В самом деле, в силу отсутствия связи между действиями отдельных абонентов вероятность любого данного числа вызовов в интервале времени (t_1, t_2) практически не зависит от от того, сколько вызовов поступает в течение других интервалов времени, не пересекающихся с интервалом (t_1, t_2). Вероятность практически одновременного поступления двух и больше вызовов можно считать равной нулю. Поэтому можно считать выполненными и условия 2) и 3). Тогда искомую вероятность можно вычислить, пользуясь распределением Пуассона, по формуле (46) при $\mu = \int_{t_1}^{t_2} \lambda(\tau)\, d\tau$.

Вместо вызовов телефонной станции можно рассматривать вызовы скорой помощи, предъявления определенных объектов техники в ремонтную организацию, выезды транспортных средств на перекресток, прибытия клиентов в обслуживающую организацию и т. п. Во всех таких случаях общее число событий, происходящих в течение длительного интервала времени, практически неограничено, а для относительно малых интервалов времени можно считать выполненными условия 1) —3), при которых поток можно считать пуассоновским, конечно приближенно, как всегда в задачах практики. То же относится и к явлениям радиоактивного распада вещества. Потоки распада атомов можно считать пуассоновскими.

9.5. Случайное распределение точек в пространстве. Распределение Пуассона получается также и в случае, когда точки распределяются случайным образом не на числовой оси, а на плоскости, поверхности или в пространстве. Предположим, что выполнены условия:

1) для любых двух непересекающихся областей вероятность попадания любого данного числа точек в одну не зависит от того, сколько точек попадает в другую;

2) вероятность попадания одной точки в бесконечно малую область представляет собой бесконечно малую того же порядка, что и площадь (объем) этой области;

3) вероятность попадания более одной точки в бесконечно малую область представляет собой бесконечно малую величину высшего порядка по сравнению с площадью (объемом) этой области.

Тогда вероятность попадания m точек в данную область B определяется формулой (46) при

$$\mu = \int_B \lambda(\tau)\, d\tau. \qquad (47)$$

Все предыдущие выкладки остаются при этом в силе. Только вероятности p_m следует рассматривать не как функции t_0 и t, а как функции величины μ и $\lambda(t)\Delta t$ во всех формулах заменить величиной $\Delta\mu$.

Пример 22. При разрыве снаряда неконтактного действия осколки распределяются по некоторой поверхности с плотностью $\lambda(t)$. При очень большом общем числе осколков условие 1) можно считать выполненным для любых областей, малых по сравнению с общей областью, покрываемой осколками. Условия 2) и 3) можно считать выполненными в силу практической невозможности попадания двух и более осколков в одну точку. Поэтому для вычисления вероятностей попадания различного числа осколков в данные области (обычно малые по сравнению с общей областью разлета осколков) можно пользоваться распределением Пуассона, вычислив μ по формуле (47).

9.6. Пуассоновское приближение биномиального распределения. Заметим, что вычисления по формуле (29), определяющей биномиальное распределение, при большом n становятся очень громоздкими. Однако при малой вероятности p биномиальное распределение при большом n хорошо аппроксимируется пуассоновским распределением. Действительно, при малой вероятности p для всех $u \in (0, 1)$

$$q + pu = 1 + p(u-1) \approx e^{p(u-1)} = e^{-p} \cdot e^{pu} \text{ *)}.$$

Подставив это выражение в формулу (28) для производящей функции, находим

$$P_{m,n} \approx \frac{(np)^m}{m!} e^{-np} \qquad (m = 0, 1, 2, \ldots). \tag{48}$$

Эта формула дает пуассоновское приближение биномиального распределения. Она достаточно точна при большом числе опытов n и при достаточно малой вероятности p события в одном опыте.

Аналогично получается приближенная формула (46) при $\mu = p_1 + \ldots + p_n$ для вероятности $P_{m,n}$ в случае переменных условий опыта при большом n и малых $p_1, \ldots, p_n$.

Предлагаем читателю самостоятельно показать, что распределение Пуассона является пределом последовательности биномиальных распределений при $p = \mu/n$ ($n = 1, 2, \ldots$), где μ — данная постоянная.

---

*) Предоставляем читателю самостоятельно убедиться в этом, представив e^z формулой Маклорена с остаточным членом второго порядка в форме Лагранжа.

Глава 2

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 1. Общие определения. Дискретные случайные величины

1.1. Определение случайной величины. В п. 1.2.2 было дано интуитивное определение случайной величины, основанное на экспериментально наблюдаемых фактах, и было показано, что с каждой случайной величиной можно связать некоторые события — ее попадания в различные множества. Для изучения случайных величин необходимо, чтобы для некоторого множества таких событий были определены вероятности, т. е. чтобы это множество событий принадлежало полю событий $\mathscr{S}$, связанному с данным опытом. Мало того, целесообразно потребовать, чтобы это множество событий само представляло собой поле событий (подполе поля $\mathscr{S}$). Таким образом, мы приходим к следующему определению случайной величины.

Случайной величиной называется величина, которая принимает в результате опыта одно из множества возможных значений и с которой связано некоторое поле событий — ее попаданий в заданные множества, — содержащееся в основном поле событий $\mathscr{S}$.

1.2. Скалярные и векторные случайные величины. Случайные величины могут быть как скалярными, так и векторными. В соответствии с общим определением вектора мы будем называть *векторной случайной величиной* или *случайным вектором* любую упорядоченную совокупность скалярных случайных величин. Так, например, n-мерный случайный вектор X представляет собой совокупность n скалярных случайных величин $\{X_1, \ldots, X_n\}$. Эти величины $X_1, \ldots, X_n$ будем называть *координатами* случайного вектора X.

Координаты случайного вектора в общем случае могут быть комплексными случайными величинами (принимающими в результате опыта комплексные числовые значения). Однако от комплексных величин всегда можно

избавиться, заменив каждую комплексную величину парой действительных — ее действительной и мнимой частями. Таким образом, n-мерный вектор с комплексными координатами всегда можно рассматривать как $2n$-мерный вектор с действительными координатами. Впрочем, это не всегда выгодно, во многих случаях удобнее рассматривать комплексные случайные величины. В дальнейшем будем для краткости называть вектор с комплексными координатами комплексным вектором, а вектор с действительными координатами действительным вектором.

Очевидно, что вместо случайного вектора можно рассматривать случайную точку в соответствующем пространстве. Это совершенно равноценные понятия.

Кроме скалярных и конечномерных векторных случайных величин, в теории вероятностей приходится изучать и случайные величины более сложной природы, в частности случайные бесконечные числовые последовательности (бесконечномерные векторы) и случайные функции. Мы будем изучать здесь только конечномерные случайные величины.

1.3. Распределение случайной величины. Каждое событие поля $\mathscr{S}_x \subset \mathscr{S}$, связанного со случайной величиной X, представляет собой попадание величины X в некоторое множество A. Это событие записывается в виде $X \in A$ (X принадлежит множеству A). Из свойств поля событий следует, что множества A, соответствующие всем событиям $\{X \in A\} \in \mathscr{S}_x$, образуют алгебру или σ-алгебру множеств $\mathscr{A}$ в пространстве значений величины X. Соответствие между множествами $A \in \mathscr{A}$ и вероятностями попадания в них величины X представляет собой распределение вероятностей в пространстве значений величины X. Это распределение вероятностей называется *распределением* (*законом распределения*) случайной величины X.

В дальнейшем будем рассматривать только распределения действительных случайных величин (скалярных или векторных). При этом всегда будем отождествлять события вида $X \in A$ с соответствующими множествами A.

В некоторых случаях все возможные значения случайной величины, за исключением, может быть, некоторого множества значений N, имеющего нулевую вероятность ($P(N) = 0$), принадлежат некоторому подпространству или какому-нибудь другому многообразию пространства ее значений. В таких случаях говорят, что распре-

деление случайной величины *полностью сосредоточено* на этом подпространстве или многообразии.

Пространство значений случайной величины X с алгеброй или σ-алгеброй $\mathcal{A}$ множеств в нем и определенной на этих множествах вероятностью образуют вероятностное пространство, называемое *вероятностным пространством случайной величины* X.

1.4. Дискретная случайная величина. Рассмотренные в п. 1.2.1 и в §§ 1.8, 1.9 примеры случайных величин показывают, что некоторые случайные величины имеют только конечное множество возможных значений, как, например, число появлений или частота события при n опытах. Другие имеют счетное множество возможных значений, как, например, число событий пуассоновского потока, происходящих в течение данного интервала времени. Третьи имеют несчетное множество возможных значений, как, например, время безотказной работы прибора, ошибки измерений, координаты точки попадания при стрельбе. Случайные величины первых двух типов во многих отношениях значительно проще, чем величины третьего типа. Поэтому их целесообразно выделить в отдельный класс.

Дискретной случайной величиной называется случайная величина с конечным или счетным множеством возможных значений.

1.5. Распределение дискретной случайной величины. Распределение дискретной случайной величины полностью определяется вероятностями всех ее возможных значений. Действительно, приняв за элементарные события возможные значения, получим конечное или счетное множество элементарных событий. Вероятности этих элементарных событий полностью определяют распределение дискретной случайной величины. Таким образом, если X — случайная величина с возможными значениями $x_1, \ldots, x_N$, то ее распределение определяется формулой

$$p_\nu = P(X = x_\nu) \qquad (\nu = 1, \ldots, N).$$

Сумма всех этих вероятностей равна единице,

$$\sum_{\nu=1}^{N} p_\nu = 1,$$

так как события $X = x_1, \ldots, X = x_N$ несовместны и образуют полную группу (случайная величина X принимает

в результате опыта какое-нибудь одно и только одно из значений $x_1, \ldots, x_N$).

Вероятностным пространством дискретной случайной величины X служит множество ее значений $\{x_1, \ldots, x_N\}$ с алгеброй всех подмножеств этих значений и вероятностью, равной для каждого подмножества сумме вероятностей значений, входящих в это подмножество.

Однако за вероятностное пространство дискретной случайной величины можно принять и все пространство, точками которого являются ее возможные значения, с σ-алгеброй всех подмножеств этого пространства и вероятностью, значение которой на каждом подмножестве равно сумме вероятностей возможных значений, содержащихся в этом подмножестве.

Пример 1. Распределение числа X появлений события при n опытах определяется формулой $P(X=m)=P_{m,n}$ $(m=0, 1, \ldots, n)$, где вероятности $P_{m,n}$ вычисляются по формулам § 1.8. В данном случае

$$N=n+1, \quad x_1=0, \quad x_2=1, \ldots, x_{N+1}=n,$$
$$p_1=P_{0,n}, \quad p_2=P_{1,n}, \ldots, \quad p_{n+1}=P_{n,n}.$$

В частности, биномиальное распределение определяется в силу (1.24) формулой $P(X=m)=C_n^m p^m q^{n-m}$ $(m=0, 1, \ldots n)$.

Пример 2. Распределение частоты Y события при n опытах определяется формулой

$$P\left(Y=\frac{m}{n}\right)=P_{m,n} \qquad (m=0, 1, \ldots, n).$$

Пример 3. *Распределение Пуассона* (в частности, распределение числа X событий пуассоновского потока, происходящих в течение данного интервала времени) определяется формулой (1.46):

$$P(X=m)=\frac{\mu^m}{m!}e^{-\mu} \qquad (m=0, 1, 2, \ldots).$$

Пример 4. Распределение r-мерного случайного вектора X, координатами которого служат числа $X_1, \ldots, X_r$ появлений несовместных событий $A_1, \ldots, A_r$, образующих полную группу, при n опытах определяется формулой (1.36):

$$P\left(\bigcap_{k=1}^{r}\{X_k=m_k\}\right)=P_{m_1,\ldots,m_r}=\frac{n!}{m_1!\,m_2!\ldots m_r!}p_1^{m_1}p_2^{m_2}\ldots p_r^{m_r},$$
$$m_1+\ldots+m_r=n.$$

Вектор X в данном случае имеет полиномиальное распределение, полностью сосредоточенное на $(r-1)$-мерной плоскости $x_1+\ldots+x_r=n$ r-мерного пространства.

§ 2. Непрерывные случайные величины. Плотность случайной величины

2.1. Плотность случайной величины. Распределение случайной величины с несчетным множеством возможных значений невозможно задать вероятностями отдельных значений. Поэтому необходим другой подход к таким случайным величинам.

В § 1.2, где мы имели в виду в основном случайные величины с несчетным множеством значений, мы видели, что распределение экспериментальных значений действительной случайной величины удобно характеризовать их относительной плотностью, которая представляет собой отношение частоты попадания в интервал к длине этого интервала. При неограниченном увеличении числа опытов частота стремится стабилизироваться около вероятности. Кроме того, длины интервалов Δx при неограниченном увеличении числа опытов можно неограниченно уменьшать. В результате гистограмма будет стремиться к некоторой кривой, ординаты которой выражают отношения вероятности к длине интервала. Таким образом, мы приходим к понятию плотности случайной величины.

Предположим, что поле событий $\mathcal{S}_x$, связанное со случайной величиной X, содержит попадания ее в любые интервалы (прямоугольники в случае векторной величины X).

Плотностью (или *плотностью вероятности*) скалярной случайной величины X называется предел отношения вероятности попадания ее значения в бесконечно малый интервал $[x, x+\Delta x)$ к длине этого интервала Δx при стягивании его в точку x:

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \leqslant X < x + \Delta x)}{\Delta x}. \qquad (1)$$

Кривая, изображающая плотность, обычно называется *кривой распределения*. Она представляет собой теоретический аналог гистограммы. На рис. 7 представлен примерный вид кривой распределения.

Если понимать неравенства между векторами как совместное выполнение таких же неравенств между соответствующими координатами векторов (например, $a < b$ как $a_k < b_k$, $k = 1, \ldots, n$), Δx в знаменателе формулы (1) как произведение $\Delta x_1 \Delta x_2 \ldots \Delta x_n$, а $\Delta x \to 0$ как $\max(\Delta x_1, \ldots, \Delta x_n) \to 0$, то формула (1) определит плотность n-мерного случайного вектора X.

Таким образом, *плотностью* случайного вектора называется предел отношения вероятности его попадания в бесконечно малый прямоугольник $[x, x+\Delta x)$ к объему этого прямоугольника Δx при стягивании его в точку x.

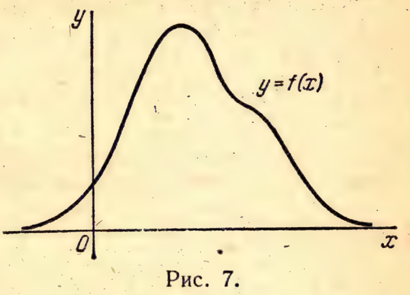


Рис. 7.

2.2. Непрерывная случайная величина. Плотность как обычная функция существует только в том случае, когда при всех x существует предел в (1) (конечный или бесконечный). Поэтому имеет смысл выделить класс случайных величин, для которых это условие выполняется.

Непрерывной случайной величиной называется такая случайная величина, вероятность попадания которой в любую бесконечно малую область бесконечно мала и для которой при каждом x существует конечный или бесконечный предел в (1) *). Из этого определения следует, что вероятность попадания непрерывной случайной величины в любую точку (или на любую кривую при $n > 1$, поверхность при $n = 3$, гиперповерхность при $n > 3$) пространства ее значений равна 0.

Если плотность $f(x)$ непрерывна в точке x, то для любой последовательности прямоугольников $\{R_p\}$, содержащих точку x, сходящейся к точке x,

$$R_{p+1} \subset R_p, \quad \lim R_p = \bigcap_{p=1}^{\infty} R_p = x, \quad f(x) = \lim \frac{P(X \in R_p)}{v(R_p)}, \qquad (2)$$

где $v(R_p)$ — объем прямоугольника R_p (длина интервала R_p при $n = 1$, площадь при $n = 2$).

► Для доказательства формулы (2) зададим произвольно малое $\varepsilon > 0$ и выберем настолько малый прямоугольник R_{p_0}, чтобы для любой точки $x' \in R_{p_0}$ было

$$|f(x') - f(x)| < \frac{\varepsilon}{2}. \qquad (3)$$

---

*) Достаточно, чтобы предел в (1) существовал почти при всех x, т. е. при всех x, кроме некоторого множества нулевой меры. *Множеством нулевой меры* называется такое множество, которое можно покрыть интервалами (прямоугольниками) сколь угодно малой суммарной длины (площади, объема).

Тогда для любого $p > p_0$ неравенство (3) будет выполнено при всех $x' \in R_p$, так как $R_p \subset R_{p_0}$ при $p > p_0$.

Пусть x_p — ближайшая к началу координат вершина прямоугольника R_p, $x_p \leqslant x'$ для всех $x' \in R_p$. На основании (3)

$$|f(x_p) - f(x)| < \frac{\varepsilon}{2}. \quad (4)$$

С другой стороны, из (1) следует, что для всех достаточно малых прямоугольников R_p

$$\left| \frac{P(X \in R_p)}{v(R_p)} - f(x_p) \right| < \frac{\varepsilon}{2}. \quad (5)$$

Из (4) и (5) следует, что для всех достаточно больших p

$$\left| \frac{P(X \in R_p)}{v(R_p)} - f(x) \right| \leqslant \left| \frac{P(X \in R_p)}{v(R_p)} - f(x_p) \right| + \\ + |f(x_p) - f(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Отсюда вследствие произвольности $\varepsilon > 0$ следует (2). ◀

Будем называть областью любое связное множество точек, имеющее определенный объем (длину при $n = 1$, площадь при $n = 2$). Легко понять, что любую область можно представить в виде счетного объединения попарно непересекающихся прямоугольников.

Пусть B — любая область. Если плотность $f(x)$ непрерывна в замкнутой области, полученной путем добавления к области B ее границы, то

$$m \leqslant \frac{P(X \in B)}{v(B)} \leqslant M, \quad (6)$$

где m и M — нижняя и верхняя грани значения плотности $f(x)$ в области B:

$$m = \inf_{x \in B} f(x), \qquad M = \sup_{x \in B} f(x).$$

▶ Пусть $\{R_p\}$ — последовательность попарно непересекающихся прямоугольников, объединением которых является область B. Предположим, что $P(X \in B)/v(B) > M$. Тогда это же неравенство справедливо, по крайней мере, для одного из прямоугольников, из которых составлена область B, скажем для R_k: $P(X \in R_k)/v(R_k) > M$. Действительно, если $P(X \in R_k)/v(R_k) < M$ для всех прямо-

угольников R_k, то в силу аксиомы сложения

$$P(X \in B) = \sum_{k=1}^{\infty} P(X \in R_k) < M \sum_{k=1}^{\infty} v(R_k) = Mv(B).$$

Обозначим прямоугольник R_k, для которого $P(X \in R_k)/v(R_k) > M$, через $R^{(1)}$. Разделим каждую сторону прямоугольника $R^{(1)}$ пополам. Тогда прямоугольник $R^{(1)}$ разобьется на 2^n равных прямоугольников и, по крайней мере, для одного из них, скажем для $R_l^{(1)} = R^{(2)}$, будет выполнено то же неравенство: $P(X \in R^{(2)})/v(R^{(2)}) > M$. Продолжая этот процесс деления прямоугольников, получим последовательность прямоугольников $\{R^{(p)}\}$, для каждого из которых

$$\frac{P(X \in R^{(p)})}{v(R^{(p)})} > M \qquad (p = 1, 2, \ldots). \tag{7}$$

Пусть x_0 — точка, принадлежащая всем прямоугольникам $R^{(p)}$. Очевидно, что такая точка существует и единственна. Следовательно, последовательность прямоугольников $\{R^{(p)}\}$ сходится к точке x_0 и в силу (2) и (7)

$$f(x_0) = \lim_{p \to \infty} \frac{P(X \in R^{(p)})}{v(R^{(p)})} > M.$$

Полученное противоречие доказывает правое неравенство (6). Аналогично доказывается левое. ◀

Из (6) следует, что формула (2) справедлива и в том случае, когда $\{R_p\}$ представляет собой произвольную последовательность областей, содержащих точку x, и сходящуюся к точке x.

▶ Действительно, обозначив через m_p и M_p нижнюю и верхнюю грани $f(x)$ в области R_p, применив неравенства (6) к каждой области R_p и приняв во внимание, что последовательности $\{m_p\}$ и $\{M_p\}$ сходятся к $f(x)$ в силу непрерывности $f(x)$, убеждаемся в справедливости высказанного утверждения. ◀

2.3. Вероятность попадания в область. Определим вероятность попадания непрерывной случайной величины X в данную область A.

▶ Предположим сначала, что плотность $f(x)$ величины X непрерывна в замкнутой ограниченной области A. Возьмем произвольную последовательность разбиений

области A

$$A=\bigcup_{k=1}^{N_p} A_k^{(p)} \qquad (p=1, 2, \ldots).$$

На рис. 8 показано одно такое разбиение для двумерного вектора X $(n=2)$. Пусть d_p — максимальный из наибольших диаметров областей $A_k^{(p)}$ $(k=1, \ldots, N_p)$:

$$d_p=\max_k \sup_{x,\, x' \in A_k^{(p)}} |x-x'|, \quad \lim_{n\to\infty} d_p=0,$$

$m_k^{(p)}$, $M_k^{(p)}$ — нижняя и верхняя грани $f(x)$ в области $A_k^{(p)}$. Тогда при всех p будем иметь:

$$P(X \in A)=\sum_{k=1}^{N_p} P(X \in A_k^{(p)})$$

и в силу (6)

$$m_k^{(p)} \leqslant P(X \in A_k^{(p)})/v(A_k^{(p)}) \leqslant M_k^{(p)}.$$

Следовательно, при всех p

$$\sum_{k=1}^{N_p} m_k^{(p)} v(A_k^{(p)}) \leqslant P(X \in A) \leqslant \sum_{k=1}^{N_p} M_k^{(p)} v(A_k^{(p)}). \qquad (8)$$

Из теории интегралов Римана известно, что при $p \to \infty$ обе суммы в (8) сходятся к одному и тому же пределу, равному интегралу от функции $f(x)$, распространенному на область A. Следовательно, вероятность попадания непрерывной случайной величины X в область A определяется формулой

$$P(X \in A)=\int_A f(x)\,dx. \qquad (9)$$

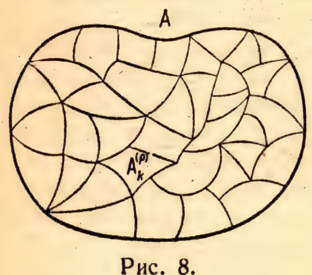


Рис. 8.

Если плотность $f(x)$ кусочно непрерывна и ограничена в замкнутой области A, то разбив A на части, в каждой из которых $f(x)$ непрерывна, и применив аксиому сложения, убеждаемся в том, что формула (9) справедлива и в случае кусочно непрерывной ограниченной плотности $f(x)$.

Если плотность $f(x)$ неограничена в области A или сама область A неограничена, то, представив A как предел последовательности ограниченных областей, в каждой из которых $f(x)$ ограничена, как это делается при определении несобственных интегралов, убеждаемся в справедливости формулы (9) и в таких случаях. При этом несобственный интеграл в (9) всегда существует как предел последовательности неотрицательных интегралов, не превосходящих 1. ◀

Таким образом, *вероятность попадания случайной величины X в данную область A равна интегралу от плотности этой величины по области A.*

В случае скалярной величины X формула (9) определяет вероятность попадания величины X в интервал $A = (\alpha, \beta)$:

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)\,dx. \tag{10}$$

В случае n-мерного вектора X интеграл в (9) представляет собой n-кратный интеграл по всем координатам вектора x. Область A обычно определяется некоторыми неравенствами, из которых находятся пределы интегрирования по всем переменным.

Величина $f(x)\,dx$, которая в силу (1) представляет собой с точностью до бесконечно малых высшего порядка вероятность попадания величины X в бесконечно малую область объема dx, называется *элементом вероятности*.

Вероятностным пространством непрерывной случайной величины X служит пространство ее значений с алгеброй множеств, для которых существует интеграл Римана в (9), и вероятностью, определяемой формулой (9) *). Таким образом, плотность непрерывной случайной величины полностью определяет ее распределение.

2.4. Свойства плотности. Изучим теперь основные свойства плотности.

1) Из определения непосредственно следует, что плотность не может быть отрицательной, $f(x) \geqslant 0$.

---

*) Легко понять, что совокупность всех множеств, для которых формула (9) и аксиома сложения определяют вероятность, образует алгебру, но не σ-алгебру. Чтобы класс множеств, для которых формула (9) определяет вероятность, был σ-алгеброй, необходимо понимать интеграл как интеграл Лебега. Тогда формула (9) определит вероятность на σ-алгебре борелевских множеств [42].

2) Учитывая, что выполнение неравенств $-\infty < X < \infty$ представляет собой достоверное событие (как в случае скалярной, так и в случае векторной величины X), из (9) получаем

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \qquad (11)$$

В случае n-мерной векторной величины X интеграл здесь следует понимать как n-кратный интеграл от $-\infty$ до ∞ по всем координатам $x_1, \ldots, x_n$ вектора x.

Таким образом, *плотность неотрицательна и интеграл от нее по всему пространству значений случайной величины равен единице.* Любая функция, обладающая этими двумя свойствами, может служить плотностью случайной величины.

Пример 5. *Равномерное распределение.* Так называется распределение непрерывной случайной величины, плотность которой постоянна в некотором интервале (a, b) и равна нулю вне этого интервала:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{при } x \in (a, b), \\ 0 & \text{при } x \notin (a, b). \end{cases} \qquad (12)$$

Равномерное распределение характерно для фазы случайных колебаний. Во многих задачах практики приходится рассматривать гармонические колебания со случайными амплитудой и фазой. В таких случаях фаза часто оказывается случайной величиной, равномерно распределенной в пределах периода колебаний.

Равномерное распределение характерно также для ошибок грубых измерений. Если измерение какой-либо величины производится с точностью до целых делений шкалы, без определения на глаз доли деления, то ошибка измерения может иметь любое значение, не превосходящее по абсолютной величине половины деления шкалы, причем нет никаких оснований считать вероятности разных значений различными. Больше того, можно с уверенностью сказать, что при большом числе таких измерений все значения ошибки в пределах от минус половины деления до плюс половины деления шкалы будут встречаться одинаково часто. Поэтому ошибка грубых измерений, производимых с точностью до целых делений шкалы, представляет собой случайную величину, равномерно распределенную в пределах от $-\Delta/2$ до $+\Delta/2$, где Δ — цена деления шкалы.

Равномерным распределением пользуются также при вычислении геометрических вероятностей на числовой прямой (см. пример 1.3).

Пример 6. *Нормальное распределение* определяется плотностью

$$f(x) = \sqrt{\frac{c}{2\pi}} \exp\left\{-\frac{1}{2} c (x-a)^2\right\}, \qquad c > 0. \qquad (13)$$

Это распределение, занимающее особое место среди всех распределений, мы изучим в § 3.6.

Пример 7. *Логарифмическое нормальное* (*логнормальное*) *распределение* определяется плотностью

$$f(x)=\sqrt{\frac{c}{2\pi}}\exp\left\{-\frac{1}{2}c(\ln x-a)^2-a-\frac{1}{2c}\right\}1(x),\qquad c>0, \tag{14}$$

где $1(x)$ — единичная ступенчатая функция, равная 1 при $x>0$ и 0 при $x<0$. Случайная величина с таким распределением неотрицательна.

Пример 8. *γ-распределение* определяется плотностью

$$f(x)=\frac{k^{\mu+1}}{\Gamma(\mu+1)}x^{\mu}e^{-kx}1(x),\qquad \mu>-1, \tag{15}$$

где $\Gamma(z)$ — *гамма-функция*, определяемая формулой

$$\Gamma(z)=\int\limits_0^{\infty}e^{-t}t^{z-1}\,dt,\qquad z>0. \tag{16}$$

Частным случаем γ-распределения при $\mu=0$ является *показательное распределение*

$$f(x)=ke^{-kx}1(x). \tag{17}$$

Показательное распределение широко применяется в теории надежности технических устройств и в теории обслуживания.

Частный случай γ-распределения при любом натуральном μ называется *распределением Эрланга*.

Для вычисления значений $f(x)$ и вероятностей попадания в интервалы $(-\infty, x)$ при разных x в случае γ-распределения с $k=1/2$ в [94] служит программа CDTR.

Пример 9. *β-распределение* определяется плотностью

$$f(x)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}1(x)1(1-x),\qquad p,q>0. \tag{18}$$

Для вычисления значений $f(x)$ и вероятностей попадания в интервалы $(-\infty, x)$ при разных x в [94] в случае β-распределения служит программа BDTR.

Использование программ BDTR и CDTR дает возможность составлять программы для вычисления вероятностей попадания в любые интервалы для величин с β- и γ-распределениями при любых значениях параметров p, q, k и μ.

Пример 10. *Равномерное распределение* в области B n-мерного пространства определяется плотностью

$$f(x)=\frac{1}{v(B)}1_B(x), \tag{19}$$

где $1_B(x)$ — индикатор области B (функция, равная 1 в области B и 0 вне области B), а $v(B)$ — объем (мера) области B.

В частности, равномерное распределение на прямоугольнике $|x|<a$, $|y|<b$ на плоскости определяется плотностью

$$f(x, y)=\frac{1}{4ab}1(a-|x|)\,1(b-|y|). \tag{20}$$

Равномерное распределение внутри эллипса $x^2/a^2+y^2/b^2=1$ на плоскости определяется плотностью

$$f(x, y)=\frac{1}{\pi ab}1\left(1-\frac{x^2}{a^2}-\frac{y^2}{b^2}\right). \tag{21}$$

Равномерным распределением на плоскости и в пространстве пользуются при вычислении геометрических вероятностей (примеры 1.4 и 1.5).

Пример 11. *Нормальное распределение* в n-мерном пространстве определяется плотностью

$$f(x)=\sqrt{\frac{|C|}{(2\pi)^n}}\exp\left\{-\frac{1}{2}(x^T-a^T)C(x-a)\right\}, \tag{22}$$

где вектор x представлен в форме матрицы-столбца, верхний индекс «T» означает операцию транспонирования матрицы, C — симметричная положительно определенная матрица, а $|C|$ — определитель матрицы C. В § 4.4 мы подробно изучим многомерное нормальное распределение.

2.5. Случайная величина как функция элементарного события. Из определения случайной величины следует, что при появлении любого элементарного события каждая случайная величина принимает одно значение. Следовательно, случайная величина представляет собой функцию элементарного события.

Так, в примере 1 число появлений X события A представляет собой функцию элементарного события, принимающую значение m при появлении любого элементарного события $\omega=\{B_1, \ldots, B_n\}$, в котором m из событий $B_1, \ldots, B_n$ совпадают с A и $n-m$ — с $\overline{A}$.

В примере 4 случайный вектор X представляет собой функцию элементарного события, принимающую значение $\{m_1, \ldots, m_r\}$ при появлении любого элементарного события $\omega=\{B_1, \ldots, B_n\}$, в котором m_k из событий $B_1, \ldots, B_n$ совпадают с A_k $(k=0, 1, \ldots, r)$ $(m_1, \ldots, m_r=0, 1, \ldots, n;$ $m_1+\ldots+m_r=n)$.

Любая случайная величина X, если принять за элементарные события точки пространства ее значений, представляет собой функцию элементарного события, которая при появлении элементарного события $\omega=x$ принимает то же значение $\omega=x$, $X=x(\omega)=\omega$.

Координата X_p любого случайного вектора $X=\{X_1, \ldots, X_n\}$ представляет собой функцию элементарного события, которая при появлении элементарного события $\omega=x=\{x_1, \ldots, x_n\}$ принимает значение $\omega_p=x_p$, $X_p=x_p(\omega)=\omega_p$ $(p=1, \ldots, n)$.

Наконец, случайная величина $Y = \varphi(X)$, представляющая собой данную функцию случайной величины X, есть функция элементарного события, которая при появлении элементарного события $\omega = x$ принимает значение $\varphi(\omega) = \varphi(x)$, $Y = \varphi(\omega)$.

Однако не всякая функция элементарного события является случайной величиной. Чтобы функция элементарного события $\varphi(\omega)$ была случайной величиной, необходимо, чтобы каждому множеству A некоторого поля в пространстве ее значений она ставила в соответствие одно определенное множество элементарных событий, принадлежащее полю $\mathcal{S}$. Такие функции называются *измеримыми* относительно соответствующих полей.

Таким образом, любая случайная величина представляет собой измеримую функцию элементарного события. Это принимается за определение случайной величины при построении теории вероятностей на основе теории меры.

§ 3. Обобщение понятия плотности

3.1. Плотность дискретной случайной величины. Предел в формуле (1), определяющей плотность, может существовать только в том случае, когда вероятность попадания случайной величины в любой бесконечно малый интервал длины (объема в случае n-мерного пространства) Δx бесконечно мала. Если вероятность в (1) представляет собой бесконечно малую низшего порядка по сравнению с Δx при каком-нибудь значении x, то предел в (1) можно считать равным бесконечности и принять $f(x) = \infty$. Если же вероятность в (1) не стремится к нулю при $\Delta x \to 0$, как, например, в случае дискретной величины X, то определить плотность как обычную функцию вообще невозможно. Определить плотность в таких случаях можно только с помощью обобщенных функций.

Чтобы подойти к определению плотности дискретной случайной величины X с возможными значениями $x_1, \ldots, x_N$ и их вероятностями $p_1, \ldots, p_N$, рассмотрим отношение в (1) при разных значениях x, предполагая, что ни один из интервалов (x_k, x_{k+1}) $(k = 1, \ldots, N-1)$ не содержит ни одного возможного значения величины X. Если точка x не совпадает ни с одной из точек $x_1, \ldots, x_N$, то при любом достаточно малом Δx интервал $[x, x + \Delta x)$ не содержит ни одного значения величины X и вероятность

попадания в этот интервал равна 0. Следовательно, плотность $f(x)$ дискретной случайной величины X существует и равна нулю при всех $x \neq x_1, \ldots, x_N$. При $x = x_k$ вероятность попадания в интервал $[x, x+\Delta x)$ равна p_k при всех достаточно малых Δx. Вследствие этого формула (1) дает $f(x) = \infty$ при $x = x_1, \ldots, x_N$. При этом вероятность попадания в интервал $(x_k - \varepsilon, x_k + \varepsilon)$ равна p_k при любом достаточно малом $\varepsilon > 0$, так как этот интервал содержит значение x_k величины X и не содержит других ее значений. С другой стороны, согласно формуле (10) вероятность попадания в этот интервал равна интегралу от плотности в пределах от $x_k - \varepsilon$ до $x_k + \varepsilon$. Следовательно, при любом достаточно малом $\varepsilon > 0$

$$\int\limits_{x_k - \varepsilon}^{x_k + \varepsilon} f(x)\,dx = p_k \qquad (k = 1, \ldots, N). \tag{23}$$

Таким образом, плотность дискретной случайной величины равна нулю всюду, кроме точек $x_1, \ldots, x_N$, в которых она «делает всплески в бесконечность», причем

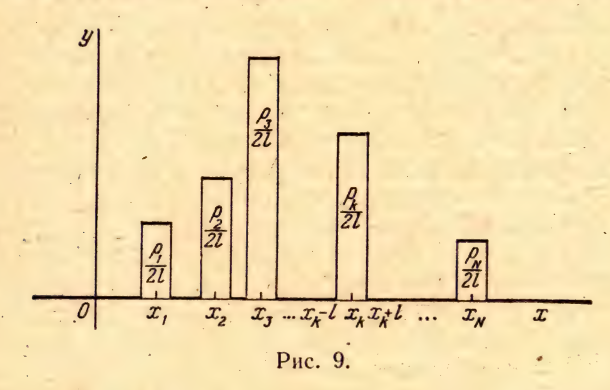


Рис. 9.

«площади всплесков» в точках $x_1, \ldots, x_N$ равны соответственно $p_1, \ldots, p_N$. Такую плотность можно представить себе как предел плотности, равной нулю всюду, кроме малых интервалов $(x_k - l, x_k + l)$ $(k = 1, \ldots, N)$, и равной соответственно $p_1/2l, \ldots, p_N/2l$ в этих интервалах (рис. 9), при $l \to 0$ *). Ясно, что этот предельный переход не опре-

*) Само собой разумеется, что вместо прямоугольных «импульсов» на рис. 9 можно взять любые другие «импульсы», имеющие основание $2l$ и площади, равные $p_1, \ldots, p_N$.

деляет плотность как функцию переменной x в обычном смысле, так как никакая функция не может удовлетворять условию (23) при любом $\varepsilon > 0$; интеграл от любой (интегрируемой) функции по бесконечно малому интервалу бесконечно мал и вследствие этого не может оставаться равным p_k при произвольно малом $\varepsilon > 0$. Тем не менее подобные предельные переходы имеют определенный смысл и целесообразны для многих физических и технических приложений. Они приводят к понятию импульсной δ-функции (см. приложение 1).

Пользуясь δ-функцией, можно определить плотность дискретной случайной величины X формулой

$$f(x) = \sum_{\nu=1}^{N} p_\nu \delta(x - x_\nu). \tag{24}$$

Действительно, подставив выражение (24) в (9), приходим к выводу, что вероятность попадания дискретной случайной величины в любую область A равна сумме вероятностей тех значений x_ν, которые содержатся в области A. Следовательно, задав плотность дискретной величины X формулой (24), мы полностью определяем распределение этой величины.

3.2. Непрерывно-дискретные случайные величины. Кроме дискретных и непрерывных, в задачах практики встречаются и такие скалярные случайные величины, для которых отличный от нуля предел в (1) существует при всех x, кроме конечного или счетного множества значений $x_1, \ldots, x_N$, имеющих отличные от нуля вероятности $p_1, \ldots, p_N$. Такие случайные величины называются *непрерывно-дискретными*.

Плотность непрерывно-дискретной скалярной случайной величины X определяется формулой

$$f(x) = f_1(x) + \sum_{\nu=1}^{N} p_\nu \delta(x - x_\nu), \tag{25}$$

где $f_1(x)$ — неотрицательная функция, представляющая собой предел в (1) при $x \neq x_1, \ldots, x_N$ *).

---

*) Множество значений случайной величины, имеющих отличные от нуля вероятности, не может быть более чем счетным. В самом деле, не больше одного значения может иметь вероятность больше 1/2, не больше трех значений могут иметь вероятности в интервале (1/4, 1/2); и вообще не больше $2^p - 1$ значений могут иметь вероятности в интервале $(2^{-p}, 2^{-p+1})$. Следовательно, расположив все такие значения в порядке убывания вероятностей, их можно все перенумеровать.

Для непрерывно-дискретной случайной величины X формула (11) дает

$$\int\limits_{-\infty}^{\infty} f_1(x)\,dx + \sum_{\nu=1}^{N} p_\nu = 1.$$

Любая функция вида (25), удовлетворяющая этому условию и условию $f_1(x) \geqslant 0$, может быть плотностью непрерывно-дискретной величины.

3.3. Непрерывно-дискретные случайные векторы. Распределение случайного вектора может иметь значительно более сложную структуру, чем распределение скалярной величины. Отличные от нуля вероятности могут быть сосредоточены не только в отдельных точках n-мерного пространства, но и на кривых, поверхностях, а при $n > 3$ и на многообразиях большей размерности. В задачах практики встречаются только такие случайные векторы, у которых во всех точках пространства значений, кроме, может быть, конечного или счетного множества многообразий с сосредоточенными вероятностями, существует предел конечный или бесконечный в (1). Любой такой вектор имеет плотность, которая может содержать δ-функции. Если на m-мерном многообразии $h(x) = 0$, $m < n$, сосредоточена отличная от нуля вероятность p, то плотность случайного вектора X содержит слагаемое вида $g(x)\,\delta(h(x))$. При интегрировании по любым $n-m$ координатам вектора x δ-функция исчезает и получается выражение $g(x)\,|J(x)|$, где согласно правилу замены переменных в кратном интеграле $J(x)$ — якобиан координат вектора $h(x)$ по соответствующим $n-m$ координатам вектора x. Интеграл от этого выражения по оставшимся m координатам вектора x равен p.

Таким образом, если в пространстве значений вектора X положительные вероятности $p_1, \ldots, p_N$ сосредоточены на многообразиях $h_1(x) = 0, \ldots, h_N(x) = 0$, а в остальных точках существует предел в (1), то плотность вектора X определяется формулой

$$f(x) = f_1(x) + \sum_{k=1}^{N} g_k(x)\,\delta(h_k(x)), \qquad (26)$$

где $f_1(x)$, $g_1(x), \ldots, g_N(x)$ — неотрицательные функции, причем чем $g_1(x), \ldots, g_N(x)$ зависят от выбора соответствующих функций $h_1(x), \ldots, h_N(x)$. В частности, уравнение k-го многообразия иногда можно записать, выра-

зив явно какие-нибудь координаты вектора x через остальные. Приняв совокупность этих координат за вектор x'', а совокупность остальных координат за вектор x', представим уравнение многообразия в виде $x''=\varphi_k(x')$. Этому соответствует функция $h_k(x)=x''-\varphi_k(x')$. При этом точку x_k, в которой сосредоточена вероятность p_k, можно считать многообразием нулевой размерности с уравнением $x=x_k$. В этом случае $h_k(x)=x-x_k$ и можно принять $g_k(x)=g_k(x_k)=p_k$.

В частном случае дискретного вектора X, все распределение которого сосредоточено в точках $x_1, \ldots, x_N$,

$$f_1(x)\equiv 0, \quad h_k(x)=x-x_k,$$
$$g_k(x)=p_k \qquad (k=1, \ldots, N)$$

и (26) принимает вид (24). В другом частном случае непрерывного вектора предел в (1) существует во всех точках пространства значений, $g_1(x)\equiv \ldots \equiv g_N(x)\equiv 0$ и $f(x)=f_1(x)$. Во всех остальных случаях вектор X будет непрерывно-дискретным.

Как мы увидим в примере 4.5, координаты непрерывно-дискретного случайного вектора могут быть непрерывными величинами.

3.4. Вырожденные распределения. Рассмотрим подробнее частный случай *вырожденного* распределения величины X, полностью сосредоточенного на одном m-мерном многообразии, $m<n$, уравнение которого $h(x)=0$ допускает явное решение относительно каких-нибудь $n-m$ координат вектора x. Представив уравнение этого многообразия в виде $x''=\varphi(x')$, получим из (26) следующую формулу для плотности вектора X *):

$$f(x)=g(x')\,\delta(x''-\varphi(x')). \tag{27}$$

Таким образом, если случайная величина X'' является определенной функцией другой случайной величины X', $X''=\varphi(X')$, то их совместная плотность (плотность составного случайного вектора $X=\{X', X''\}$) определяется формулой (27). Функция $g(x')$ в (27), как мы увидим в примере 4.5, представляет собой плотность величины X'.

В частном случае вырожденного распределения величины X, сосредоточенного на подпространстве или сдви-

---

*) Так как δ-функция в (27) равна нулю всюду вне многообразия $x''=\varphi(x')$, то функцию $g(x)=g(x', x'')$ можно заменить ее значением на этом многообразии: $g(x', x'')=g(x', \varphi(x'))=g_1(x')$. Отбрасывая индекс у найденной таким путем новой функции, получим (27).

нутом подпространстве, имеющем уравнение $x'' = Ax' + c$, где A — в общем случае прямоугольная матрица, а c — вектор сдвига, формула (27) принимает вид

$$f(x) = g(x')\,\delta(x'' - Ax' - c). \tag{28}$$

3.5. Вероятность попадания в область. Дискретными, непрерывными и непрерывно-дискретными случайными величинами исчерпывается все множество случайных величин, встречающихся в задачах практики. Мы показали, что для этих трех типов случайных величин распределение полностью определяется плотностью. Это дает основание в приложениях теории вероятностей принять плотность за основную характеристику распределения любой конечномерной случайной величины. При этом будет справедлива формула (9) для вероятности попадания случайной величины в любую данную область. Однако в общем случае при интегрировании в формуле (9) следует учитывать все принадлежащие области A части ее границы, на которых сосредоточена отличная от нуля вероятность (на которых аргументы δ-функций, входящих в $f(x)$, обращаются в нуль). Так, например, при вычислении вероятности попадания в интервал $[\alpha, \beta)$ следует учитывать, что вероятность значения α случайной величины X может быть отличной от нуля. Тогда получим

$$P(\alpha \leqslant X < \beta) = \int\limits_{\alpha-0}^{\beta-0} f(x)\,dx = \lim_{\varepsilon \downarrow 0} \int\limits_{\alpha-\varepsilon}^{\beta-\varepsilon} f(x)\,dx.$$

§ 4. Функция распределения

4.1. Функция распределения и ее связь с плотностью. *Функцией распределения* случайной величины X называется вероятность неравенства $X < x$, рассматриваемая как функция параметра x:

$$F(x) = P(X < x). \tag{29}$$

Функция распределения непрерывной, дискретной или непрерывно-дискретной случайной величины может быть выражена через плотность. Для этого достаточно вычислить вероятность неравенства $X < x$ в формуле (29) по формуле (9) или (10):

$$F(x) = P(-\infty < X < x) = \int\limits_{-\infty}^{x} f(u)\,du, \tag{30}$$

где в n-мерном случае интеграл представляет собой n-кратный интеграл, причем интегрирование по переменной u_i производится от $-\infty$ до x_i $(i = 1, \ldots, n)$.

Дифференцируя формулу (30) по x в случае скалярной величины X и вспоминая, что производная интеграла по верхнему пределу равна значению подынтегральной функции при этом верхнем пределе, получаем

$$f(x) = F'(x). \tag{31}$$

Таким образом, *плотность скалярной случайной величины есть производная функции распределения.*

В случае векторной величины X напишем формулу (30) в скалярной форме

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(u_1, \ldots, u_n)\, du_1 \ldots du_n$$

и продифференцируем ее один раз по x_1, один раз по x_2, и т. д., один раз по x_n. В результате получим

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}. \tag{32}$$

Таким образом, *плотность n-мерной случайной величины представляет собой симметричную смешанную производную n-го порядка функции распределения.*

4.2. Свойства функции распределения скалярной величины. Изучим свойства функции распределения скалярной случайной величины.

1) Из формулы (30) следует, что *функция распределения — неубывающая функция x.*

2) Из (30) и свойства (11) плотности следует, что

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, \quad F(\infty) = \lim_{x \to \infty} F(x) = 1.$$

3) Из (30) и соглашения в п. 2.1 о том, как надо понимать интегралы от плотности, следует, что *функция распределения непрерывна слева: $F(x) = F(x-0)$.*

4) Функция распределения дискретной случайной величины возрастает скачками в точках $x_1, \ldots, x_N$ соответственно на $p_1, \ldots, p_N$ и постоянна в любом интервале, не содержащем ни одного из значений $x_1, \ldots, x_N$, так как вероятность события $X < x$ не изменяется при изменении x в таком интервале. Таким образом, функция распределения дискретной случайной величины изображается ступенчатой линией (рис. 10).

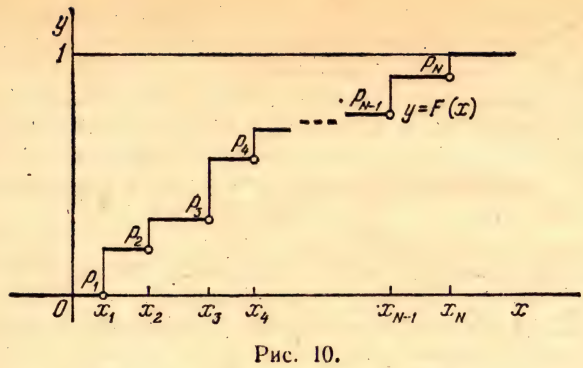


Рис. 10.

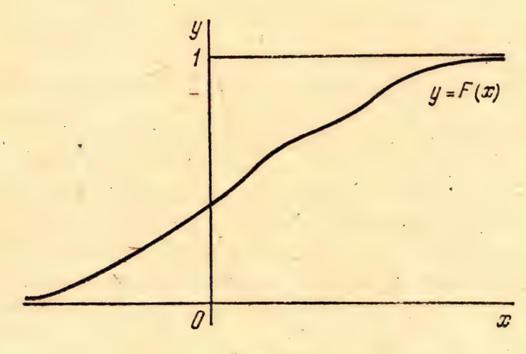


Рис. 11.

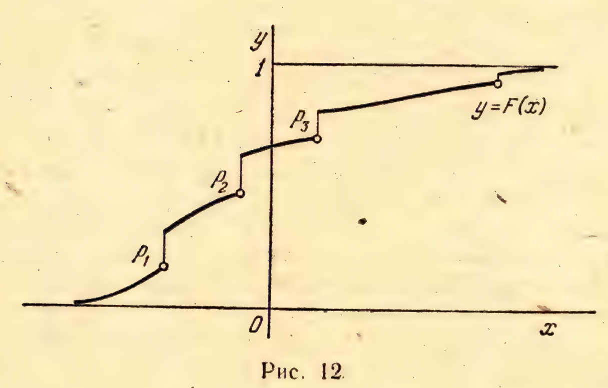


Рис. 12.

5) Из свойств интеграла с переменным верхним пределом следует, что функция распределения непрерывной случайной величины непрерывна и дифференцируема на всей числовой оси. Однако ее производная может быть и разрывной функцией, как, например, в случае равномерного или показательного распределения (рис. 11).

6) Функция распределения непрерывно-дискретной случайной величины имеет точки разрыва $x_1, \ldots, x_N$ со скачками соответственно $p_1, \ldots, p_N$ и непрерывна и дифференцируема во всех остальных точках числовой оси (рис. 12).

4.3. Вероятность попадания скалярной величины в интервал. Зная функцию распределения скалярной случайной величины, можно вычислить вероятность ее попадания в любой полузамкнутый интервал $[\alpha, \beta)$.

► В самом деле, так как события $X < \alpha$ и $\alpha \leqslant X < \beta$ несовместны и их объединение есть событие $X < \beta$, то

$$P(X < \beta) = P(X < \alpha) + P(\alpha \leqslant X < \beta),$$

откуда в силу (29) следует

$$P(\alpha \leqslant X < \beta) = F(\beta) - F(\alpha). \blacktriangleleft \qquad (33)$$

Таким образом, *вероятность попадания случайной величины в данный интервал равна приращению ее функции распределения на этом интервале.*

Положив в (33) $\beta = \alpha + \varepsilon$, $\varepsilon > 0$, и устремив ε к нулю, получаем

$$P(X = \alpha) = F(\alpha + 0) - F(\alpha). \qquad (34)$$

Таким образом, *вероятность попадания случайной величины в данную точку равна скачку ее функции распределения в данной точке.*

Пример 12. Найти закон распределения времени T безотказной работы системы с момента начала эксплуатации при постоянной интенсивности отказов λ (см. пример 1.11).

Найдем сначала функцию распределения этой случайной величины, т. е. вероятность того, что случайная величина T будет меньше, чем t: $F(t) = P(T < t)$. Но это есть вероятность того, что система откажет до момента t. Вероятность безотказной работы системы до момента t, т. е. $P(T > t)$, представляет собой функцию надежности системы $p(t)$. При постоянной интенсивности отказов λ функция надежности $p(t) = e^{-\lambda t}$. Следовательно,

$$F(t) = P(T < t) = 1 - P(T \geqslant t) = 1 - p(t) = (1 - e^{-\lambda t})1(t).$$

При $t < 0$ эта вероятность равна 0, потому что исправная система не может отказать до начала ее эксплуатации (конечно, предполагается, что ее начинают эксплуатировать в исправном состоянии).

Следовательно, $F(t)=0$ при $t<0$. Дифференцируя полученную формулу, находим плотность времени T безотказной работы

$$f(t)=\lambda e^{-\lambda t}1(t).$$

Таким образом, время безотказной работы системы (момент первого отказа) при постоянной интенсивности отказов λ имеет показательное распределение с параметром $k=\lambda$.

Пример 13. Найдем распределение интервала времени T между двумя последовательными событиями пуассоновского потока событий постоянной интенсивности λ. В этом случае среднее число событий, происходящих в интервале времени длительности t, равно $\mu=\lambda t$. Функция распределения интервала T между двумя последовательными событиями потока есть, очевидно, вероятность того, что в интервале времени длительности t произойдет хотя бы одно событие:

$$F(t)=P(T<t).$$

Вероятность противоположного события — в интервале времени длительности t не произойдет ни одного события — вычисляется по формуле (1.46) при $\mu=\lambda t$, $m=0$:

$$p_0=e^{-\lambda t}.$$

Следовательно,

$$F(t)=1-p_0=1-e^{-\lambda t}.$$

Но это — та же формула, которая была получена в предыдущем примере для функции распределения времени безотказной работы системы. Таким образом, интервал времени между двумя последовательными событиями в пуассоновском потоке постоянной интенсивности представляет собой случайную величину, распределенную по показательному закону.

Пример 14. Некоторая система эксплуатируется в течение времени t_0. Если она откажет в течение этого времени, то ее ремонтируют и снова используют, пока она не прослужит срок t_0. Найти закон распределения времени S работы системы после первого ремонта.

Очевидно, что случайная величина S связана с временем T работы системы до первого отказа соотношениями

$$S=0 \text{ при } T\geqslant t_0, \qquad S=t_0-T \text{ при } T<t_0.$$

Так как $S\geqslant 0$, то при $s\leqslant 0$ ее функция распределения $G(s)$ равна 0. При $s>0$ она выражается через функцию распределения $F(t)$ времени T безотказной работы системы формулой

$$G(s)=P(S<s)=P(t_0-T<s)=P(T>t_0-s)=$$
$$=1-P(T<t_0-s)=1-F(t_0-s). \qquad \text{(I)}$$

Подставляя сюда выражение $F(t)$ из примера 12, получаем

$$G(s)=e^{-\lambda(t_0-s)} \text{ при } 0<s<t_0. \qquad \text{(II)}$$

Наконец, $G(s)=1$ при $s>t_0$, так как $S<t_0$. Из (II) видно, что

$$G(+0)=e^{-\lambda t_0}.$$

Таким образом, функция $G(s)$ имеет разрыв со скачком $e^{-\lambda t_0}$ в точке $s=0$. Во всех остальных точках числовой оси она непре-

рывна. При любом $s \in (0, t_0)$ она имеет производную $\lambda e^{-\lambda (t_0 - s)}$, а при любых $s < 0$ и $s > t_0$ ее производная равна нулю. Следовательно, S представляет собой непрерывно-дискретную случайную величину с одним исключительным значением. Ее плотность определяется формулой

$$g(s) = G'(s) = \begin{cases} \lambda e^{-\lambda (t_0 - s)} + e^{-\lambda t_0} \delta(s) & \text{при } s \in [0, t_0], \\ 0 & \text{при } s \notin [0, t_0]. \end{cases}$$

Пример 15. Текущее значение входного сигнала X нелинейного элемента представляет собой непрерывную случайную величину с функцией распределения $F(x)$. Зависимость текущего значения выходного сигнала Y этого элемента связано со значением входного сигнала X в тот же момент времени соотношением

$$Y = \begin{cases} X & \text{при } |X| \leqslant a, \\ a & \text{при } X > a, \\ -a & \text{при } X < -a \end{cases}$$

(элемент с такой характеристикой называется *ограничителем*). Найти распределение выходного сигнала.

Ясно, что функция распределения $G(y)$ случайной величины Y равна 0 при любом $y \leqslant -a$ (неравенство $Y < y$ при $y \leqslant -a$ невозможно) и равна 1 при любом $y > a$ (событие $Y < y$ при $y > a$ достоверно). При $|y| < a$ событие $Y < y$ совпадает с событием $X < y$. Следовательно, $G(y) = F(y)$ при $|y| < a$. Очевидно, что если $F(-a) > 0$, $F(a) < 1$, то функция распределения $G(y)$ имеет разрыв со скачком $F(-a)$ при $y = -a$ и разрыв со скачком $1 - F(a)$ при $y = a$. При всех остальных значениях y функция $G(y)$ непрерывна и дифференцируема. Следовательно, в этом случае Y представляет собой непрерывно-дискретную случайную величину с двумя исключительными значениями $-a$ и a. Ее плотность определяется формулой

$$g(y) = f(y) + F(-a)\,\delta(y + a) + [1 - F(a)]\,\delta(y - a) \text{ при } |y| \leqslant a,$$

где $f(x)$ — плотность случайной величины X. Вне интервала $(-a, a)$ $g(y) = 0$.

Пример 16. Пусть X — непрерывная случайная величина с плотностью $f(x)$. Найти закон распределения длины пересечения случайного интервала $(X - l, X + l)$ с данным интервалом (a, b), предполагая, что $2l < b - a$. Примером практической задачи такого типа может служить, например, следующая задача из теории стрельбы. Войска противника сосредоточены на участке линии фронта от точки a до точки b. При выстреле снаряд уничтожает на линии фронта все на интервале длины $2l$. При этих условиях снаряд поражает войска противника, расположенные на той части интервала (a, b), которая накрывается случайным интервалом $(X - l, X + l)$, где X — случайная координата центра зоны поражения. Подобные задачи о перекрытиях часто возникают в теории стрельбы.

В зависимости от значения x величины X накрытая часть интервала (a, b) будет равна или $2l$ (рис. 13, *а*), или нулю (рис. 13, *б*), или любому промежуточному значению в интервале $(0, 2l)$ (рис. 13, *в* и 13, *г*):

$$\begin{aligned} U &= 0 && \text{при } X < a - l \text{ и при } X > b + l, \\ U &= X - a + l && \text{при } a - l < X < a + l, \\ U &= b + l - X && \text{при } b - l < X < b + l, \\ U &= 2l && \text{при } a + l < X < b - l. \end{aligned}$$

Для выполнения неравенства $U < u$ при $0 < u \leqslant 2l$ необходимо, чтобы удовлетворилось одно из двух несовместных неравенств

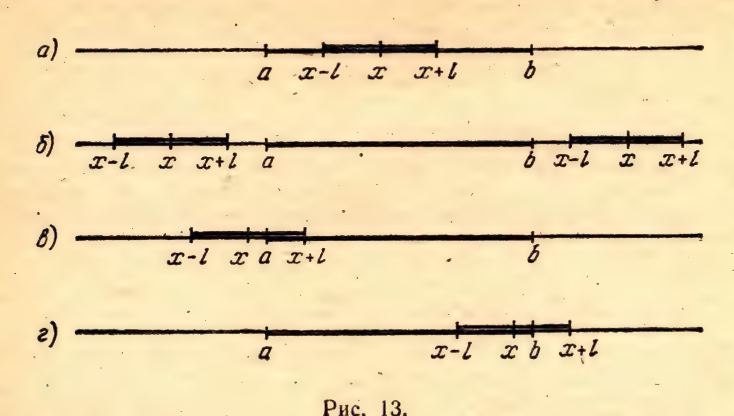


Рис. 13.

$X - a + l < u$ или $b + l - X < u$. Поэтому функция распределения $G(u)$ случайной величины U определяется формулой

$$G(u) = P(U < u) = P(X - a + l < u) + P(b + l - X < u) =$$
$$= P(X < u + a - l) + P(X > b + l - u)$$

Выразив последние вероятности через плотность по формуле (10), получим

$$G(u) = \int_{-\infty}^{u+a-l} f(x)\,dx + \int_{b+l-u}^{\infty} f(x)\,dx, \tag{I}$$

$$G(u) = 1 - \int_{u+a-l}^{b+l-u} f(x)\,dx. \tag{II}$$

Так как накрытая часть интервала (a, b) не может быть отрицательной, то $G(u) = 0$ при $u \leqslant 0$. А вследствие того, что U не может превосходить $2l$, событие $U < u$ достоверно при любом $u > 2l$, вследствие чего $G(u) = 1$ при $u > 2l$. Но из формулы (II) следует, что

$$\lim_{u \to 0} G(u) = 1 - \int_{a-l}^{b+l} f(x)\,dx,$$

$$G(2l) = 1 - \int_{a+l}^{b-l} f(x)\,dx.$$

Следовательно, функция распределения $G(u)$ терпит разрывы в точках $u=0$ и $u=2l$ со скачками, равными соответственно

$$p_1 = \lim_{u \to 0} G(u) = 1 - \int_{a-l}^{b+l} f(x)\,dx,$$

$$p_2 = 1 - G(2l) = 1 - \int_{a+l}^{b-l} f(x)\,dx.$$

Таким образом, накрытая часть интервала (a, b) представляет собой непрерывно-дискретную случайную величину U, реализации которой сплошь заполняют интервал $[0, 2l]$, и при этом вероятности двух исключительных реализаций 0 и $2l$ отличны от нуля.

Плотность накрытой части интервала (a, b) на основании (31) определяется формулой

$$g(u) = g_1(u) + p_1\delta(u) + p_2\delta(u-2l), \tag{III}$$

где $g_1(u)=0$ при $u \leqslant 0$ и при $u \geqslant 2l$, а в интервале $(0, 2l)$ равна производной функции распределения $G(u)$, определяемой формулой (I) и (II). Дифференцируя формулу (I) по правилу дифференцирования определенного интеграла по верхнему и нижнему пределам, получим при $0 < u < 2l$

$$g_1(u) = G'(u) = f(u+a-l) + f(b+l-u). \tag{IV}$$

Естественно возникает вопрос: существуют ли случайные величины, не относящиеся ни к одному из рассмотренных трех классов случайных величин, с которыми приходится встречаться в задачах практики? Чтобы убедиться в том, что существуют, достаточно привести пример такой случайной величины.

Пример 17. Рассмотрим случайную величину X, функция распределения которой $F(x)$ определяется следующим образом: принимается $F(x)=0$ при $x \leqslant 0$, $F(x)=1$ при $x \geqslant 1$; затем интервал $(0, 1)$ делится на три равные части и в средней части (при $x \in (1/3, 2/3]$) принимается $F(x) = 1/2$; каждая из оставшихся частей снова делится на три равные части, и в средних частях принимается соответственно $F(x) = 1/4$ и $F(x) = 3/4$; каждая из оставшихся частей в свою очередь, делится на 3 равные части, и на средних частях $F(x)$ принимается равной соответственно 1/8, 3/8, 5/8, 7/8; этот процесс деления оставшихся частей на 3 равные части продолжается неограниченно; и каждый раз на средней части $F(x)$ определяется как среднее арифметическое ее значений на соседних интервалах, на которых она уже определена (рис. 14). При этом в силу свойства непрерывности функции распределения слева $F(x)$ определится и во всех точках интервала $(0, 1)$, не принадлежащих ни одному из интервалов, на которых она определена описанным образом. Ясно, что построенная таким путем функция распределения $F(x)$ непрерывна. Суммарная длина интервалов, на каждом из которых она постоянна, равна 1. Действительно, один интервал постоянства $F(x)$ имеет длину 1/3, два имеют длину 1/9, четыре — 1/27, и вообще 2^p интервалов постоянства $F(x)$

имеют длину $1/3^{p+1}$. Следовательно, суммарная длина всех интервалов постоянства $F(x)$ равна

$$\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\ldots+\frac{2^p}{3^{p+1}}=\frac{1}{3}\sum_{p=0}^{\infty}\left(\frac{2}{3}\right)^p=1.$$

Таким образом, суммарная длина интервалов постоянства $F(x)$ совпадает с длиной интервала $(0, 1)$, на котором она изменяется от 0 до 1. Следовательно, множество точек роста функции $F(x)$ имеет

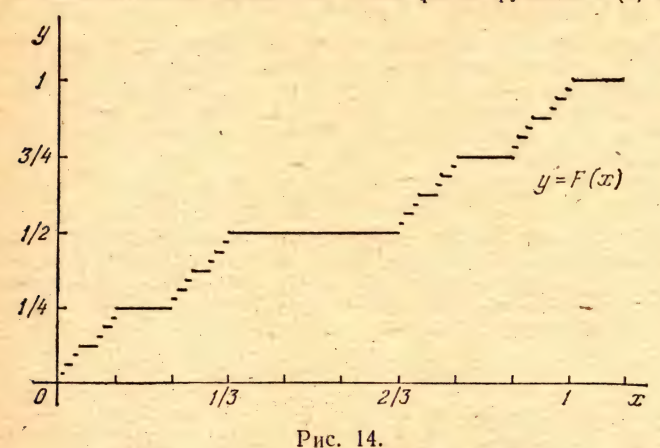


Рис. 14.

нулевую меру *). Несмотря на это, множество точек роста функции $F(x)$ несчетно. В самом деле, функция $F(x)$ непрерывна, вследствие чего ее изменение в любой точке (не на интервале, а в одной точке) равно нулю, и если бы множество ее точек роста было счетным, то она вообще не могла бы изменяться и была бы постоянной (сумма счетного множества нулей всегда равна нулю как сумма бесконечного ряда, все члены которого равны нулю).

Таким образом, функция $F(x)$ непрерывна и имеет производную, равную нулю, во всех точках числовой оси, кроме несчетного множества точек ее роста, имеющего меру нуль. В точках же роста функция $F(x)$ не имеет производной. Очевидно, что случайная величина с такой функцией распределения не является ни дискретной, ни непрерывной, ни непрерывно-дискретной.

Чтобы дать более общий пример, обозначим только что определенную функцию распределения через $F_0(x)$. Пусть $\varphi(x)$ — монотонно возрастающая функция с произвольной областью определения на числовой оси и с областью значений $[0, 1]$, $\psi(x)$ — монотонно возрастающая функция, отображающая интервал $[0, 1]$ на себя (т. е. с областью определения и областью значений $[0, 1]$). Очевидно, что случайная величина с функцией распределения $F(x)=\psi(F_0(\varphi(x)))$ также не будет ни дискретной, ни непрерывной, ни непрерывно-дискретной при любом выборе монотонно возрастающих функций φ и ψ.

---

*) См. сноску на стр. 63.

В дальнейшем будем рассматривать только такие случайные величины, которые имеют плотность, возможно содержащую линейную комбинацию δ-функций.

4.4. Вероятность попадания случайного вектора в прямоугольник. Пусть X — n-мерный случайный вектор, $F(x)$ — его функция распределения. Обозначим через $\Delta_I^{(k)}F$ приращение функции распределения $F(x)$ при изменении k-й координаты x_k вектора x и при неизменных значениях x_i остальных координат, $i \neq k$. На основании аксиомы сложения величина $\Delta_I^{(k)}F$ представляет собой вероятность попадания величины X в прямоугольник со стороной I по k-й оси и с бесконечными в обоих направлениях сторонами по остальным осям.

Применяя последовательно аксиому сложения, приходим к заключению, что вероятность попадания случайного вектора X в прямоугольник R, сторонами которого по осям $x_{k_1}, \ldots, x_{k_p}$ служат интервалы $I_1, \ldots, I_p$, а остальные стороны которого совпадают с соответствующими координатными осями, определяется формулой

$$P(X \in R) = \Delta_{I_1}^{(k_1)} \ldots \Delta_{I_p}^{(k_p)} F. \qquad (35)$$

4.5. Свойства функции распределения случайного вектора. Изучим основные свойства функции распределения случайного вектора.

1) Из определения (29) и неотрицательности плотности следует, что *функция распределения $F(x)$ случайного вектора X является неубывающей функцией каждой из координат вектора x.*

2) Из (29) и свойства (11) плотности следует, что $F(x) \to 0$, когда хотя бы одна координата вектора x стремится к $-\infty$, и $F(x) \to 1$, когда все координаты вектора x стремятся к $+\infty$.

3) Из (29) и соглашения в п. 3.5 о том, как надо понимать интегралы от плотности, следует, что *функция распределения $F(x)$ непрерывна слева по каждой координате вектора x.*

4) Из формулы (35) следует, что

$$\Delta_{I_1}^{(k_1)} \ldots \Delta_{I_p}^{(k_p)} F \geqslant 0$$

для любых натуральных $p \leqslant n$, любых $k_1, \ldots, k_p, 1 \leqslant k_1 < k_2 < \ldots < k_p \leqslant n$, и любых интервалов $I_1, \ldots, I_p$.

Любая функция $F(x)$, обладающая этими четырьмя свойствами (точнее, свойствами 2), 3) и 4), поскольку

1) представляет собой следствие свойства 4)), может быть функцией распределения случайной величины.

Это доказывается совершенно элементарно, если существует производная $F'(x)$ ($F'(x) = \partial^n F(x)/\partial x_1 \ldots \partial x_n$ в случае n-мерного вектора X), возможно содержащая линейную комбинацию δ-функций. В этом случае по доказанному в §§ 2 и 3 существует случайная величина X, для которой $f(x) = F'(x)$ служит плотностью. Функция $F(x)$ служит для этой величины X функцией распределения.

Доказательство в общем случае требует привлечения теории меры и здесь не приводится.

Глава 3

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 1. Математическое ожидание

1.1. Математическое ожидание дискретной случайной величины. Распределение полностью характеризует случайную величину. Чтобы определить распределение случайной величины, надо задать ее плотность вероятности или функцию распределения. Однако такая полная, исчерпывающая характеристика случайной величины довольно сложна. Между тем для решения многих задач практически вовсе не нужно знать распределения случайных величин, а достаточно знать лишь некоторые числа, характеризующие их распределения, так называемые *числовые характеристики* случайных величин. Для грубой характеристики скалярной случайной величины можно, так же как в § 1.2, ограничиться ее средним значением и величиной разброса возможных значений.

Чтобы подойти к определению среднего значения случайной величины по данному ее распределению, рассмотрим выборочное среднее дискретной случайной величины X с возможными значениями $x_1, \ldots, x_N$ и вероятностями этих значений $p_1, \ldots, p_N$ соответственно. Очевидно, что при достаточно большом числе опытов n каждое из значений $x_1, \ldots, x_N$ должно появляться многократно. Предположим, что величина X приняла m_1 раз значение x_1, m_2 раз значение x_2, и т. д., m_N раз значение x_N, $m_1 + \ldots + m_N = n$. Тогда выборочное среднее выразится формулой

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{N} m_k x_k = \sum_{k=1}^{N} x_k \frac{m_k}{n}.$$

Но m_k/n есть частота значения x_k $(k = 1, \ldots, N)$. Следовательно, выборочное среднее дискретной случайной величины равно сумме всех ее возможных значений, умноженных на их частоты.

При неограниченном увеличении числа опытов n частоты значений $x_1, \ldots, x_N$ стремятся стабилизироваться около их вероятностей. Поэтому, чтобы определить среднее значение случайной величины X как абстрактную характеристику ее распределения, достаточно заменить в полученной формуле частоты $m_1/n, \ldots, m_N/n$ значений $x_1, \ldots, x_N$ их вероятностями $p_1, \ldots, p_N$. Это приводит к следующему определению.

Математическим ожиданием или *средним значением* дискретной случайной величины называется сумма всех ее значений, умноженных на их вероятности.

Математическое ожидание случайной величины X мы будем обозначать MX или, сокращенно, m_x *).

На основании данного определения математическое ожидание дискретной случайной величины X выражается формулой

$$m_x = MX = \sum_{k=1}^{N} x_k p_k. \tag{1}$$

Математическое ожидание произвольной (однозначной) функции $\varphi(X)$ дискретной случайной величины X на основании данного определения выражается формулой

$$M\varphi(X) = \sum_{k=1}^{N} \varphi(x_k) p_k. \tag{2}$$

Действительно, в данном случае величина $Y = \varphi(X)$ представляет собой дискретную случайную величину с возможными значениями $\varphi(x_1), \ldots, \varphi(x_N)$ и вероятностями этих значений, равными $p_1, \ldots, p_N$ соответственно. Если некоторые из значений $\varphi(x_k)$ функции $\varphi(x)$ совпадают, то вероятность этого общего значения величины $Y = \varphi(X)$ равна сумме соответствующих вероятностей p_k. Таким образом, формула (2) определяет математическое ожидание величины $Y = \varphi(X)$ и в этом случае.

1.2. Общее определение математического ожидания. Пользуясь формулой (2.24) для плотности дискретной случайной величины, можно записать (2) в виде

$$M\varphi(X) = \int_{-\infty}^{\infty} \varphi(x) f(x)\, dx. \tag{3}$$

*) Иногда, особенно в иностранной литературе, математическое ожидание величины X обозначается EX от английского expected value (ожидаемое значение).

Эту формулу естественно положить в основу определения математического ожидания для случайных величин всех трех типов, встречающихся в задачах практики.

Математическим ожиданием функции $\varphi(X)$ случайной величины X с плотностью $f(x)$ называется интеграл (3).

В частном случае непрерывной величины X и кусочно непрерывной функции $\varphi(x)$ интеграл (3) представляет собой интеграл Римана.

Ясно, что математическое ожидание функции $\varphi(X)$ существует тогда и только тогда, когда

$$\int_{-\infty}^{\infty} |\varphi(x)| f(x)\, dx < \infty \text{ *)}. \tag{4}$$

Чтобы получить формулу для вычисления математического ожидания функции $\varphi(X)$ в случае непрерывно-дискретной величины X, подставим в (3) выражение (2.25) ее плотности.

Тогда получим

$$M\varphi(X) = \int_{-\infty}^{\infty} \varphi(x) f_1(x)\, dx + \sum_{k=1}^{N} \varphi(x_k)\, p_k. \tag{5}$$

Если в частном случае областью возможных значений величины X служит не вся числовая ось, а только какой-нибудь ее интервал (a, b), то плотность $f(x)$ равна нулю вне этого интервала и интегрирование в (3) производится от $x=a$ до $x=b$. В этом случае интеграл (4), а следовательно, и математическое ожидание величины $\varphi(X)$ заведомо существует.

---

*) Если интеграл (4) расходится, то расходится по крайней мере один из интегралов

$$\int_{-\infty}^{\infty} \varphi_-(x) f(x)\, dx, \qquad \int_{-\infty}^{\infty} \varphi_+(x) f(x)\, dx,$$

где $\varphi_-(x) = \min\{0, \varphi(x)\}$; $\varphi_+(x) = \max\{0, \varphi(x)\}$. Если расходится один из них, то $M\varphi(X) = -\infty$ или $M\varphi(x) = +\infty$. Если расходятся оба, то

$$\int_{-a}^{b} \varphi(x) f(x)\, dx$$

не стремится ни к какому пределу, когда a и b неограниченно возрастают независимо одно от другого.

В частном случае при $\varphi(x)=x$ формула (3) определяет математическое ожидание случайной величины X:

$$m_x = MX = \int_{-\infty}^{\infty} x f(x)\,dx. \tag{6}$$

В случае векторной величины X интегралы в (3) и (6) представляют собой кратные интегралы, распространенные на все пространство значений величины X. Функция $\varphi(x)$ в (3) может быть скалярной (действительной или комплексной), векторной или матричной. При этом, как обычно, интеграл от векторной (матричной) функции понимается как вектор (матрица), координатами которого (элементами которой) служат интегралы от соответствующих координат векторной (элементов матричной) функции.

Принимая во внимание сказанное, приходим к выводу, что математическое ожидание m_x случайного вектора X с координатами $X_1, \ldots, X_n$ представляет собой вектор с координатами $m_1 = MX_1, \ldots, m_n = MX_n$.

1.3. Свойства математических ожиданий. Основные свойства математического ожидания вытекают непосредственно из свойств интеграла.

1) Если функция $\varphi(x)$ представляет собой постоянную $\varphi(x)=c$, то в силу свойства (2.11) плотности вероятности

$$Mc = \int_{-\infty}^{\infty} cf(x)\,dx = c\int_{-\infty}^{\infty} f(x)\,dx = c. \tag{7}$$

Таким образом, *математическое ожидание неслучайной величины равно самой величине.*

2) Далее, из свойств интеграла следует

$$M\sum_{k=1}^{n} c_k\varphi_k(X) = \int_{-\infty}^{\infty}\sum_{k=1}^{n} c_k\varphi_k(x) f(x)\,dx = \\ = \sum_{k=1}^{n} c_k \int_{-\infty}^{\infty} \varphi_k(x) f(x)\,dx = \sum_{k=1}^{n} c_k M\varphi_k(X). \tag{8}$$

Заметим теперь, что любые комплексные случайные величины $X_k = X'_k + iX''_k$ $(k=1, \ldots, n)$ можно рассматривать как функции $2n$-мерного случайного вектора X с координатами $X'_1, X''_1, \ldots, X'_n, X''_n$, $X_k = \varphi(X)$

($k=1, \ldots, n$)*). Поэтому из (8) следует, в частности, что для любых скалярных случайных величин $X_1, \ldots, X_n$, действительных или комплексных,

$$M \sum_{k=1}^{n} c_k X_k = \sum_{k=1}^{n} c_k M X_k. \tag{9}$$

Таким образом, *математическое ожидание линейной функции случайных величин равно той же самой функции от их математических ожиданий* (свойство линейности операции математического ожидания **)).

В частном случае $n=1$ отсюда следует, что *при умножении случайной величины на число ее математическое ожидание умножается на то же число.*

3) При $c_1 = \ldots = c_n = 1$ из (9) вытекает теорема сложения математических ожиданий: *математическое ожидание суммы случайных величин всегда равно сумме их математических ожиданий*:

$$M \sum_{k=1}^{n} X_k = \sum_{k=1}^{n} M X_k. \tag{10}$$

4) Ясно, что формулы (7)—(10) справедливы как для скалярных величин, так и для матричных (в частном случае—векторных, если пользоваться представлением векторов в виде матриц-столбцов) величин c, X, $c_1, \ldots, c_n$, $X_1, \ldots, X_n$. Из (9) следует, что для любой неслучайной матрицы C и случайной векторной или матричной величины X

$$MCX = CMX. \tag{11}$$

Точно так же из (9) вытекает формула

$$MXC = (MX)C. \tag{12}$$

Само собой разумеется, что в случае прямоугольных (не квадратных) матриц C и X буква C в (11) и (12) не может означать одну и ту же матрицу.

Комбинируя (7) и (11), получаем

$$M(AX + a_0) = AMX + a_0. \tag{13}$$

---

*) В § 4.1 будет показано, что, рассматривая действительную случайную величину как координату некоторого случайного вектора и вычисляя ее математическое ожидание как математическое ожидание функции этого случайного вектора по формуле (6), мы получим тот же результат, что и при вычислении по формуле (3).

**) *Операцией* или *оператором математического ожидания* мы называем переход от случайной величины к ее математическому ожиданию.

§ 2. Моменты второго порядка

2.1. Дисперсия. Теперь можно определить абстрактные аналоги всех выборочных числовых характеристик § 1.2. Для этого достаточно заменить все выборочные средние математическими ожиданиями. Предварительно введем понятие центрированной случайной величины.

Центрированной случайной величиной называется разность между случайной величиной и ее математическим ожиданием, т. е. отклонение случайной величины от ее математического ожидания. Мы будем отмечать центрированные случайные величины индексом 0 сверху, $X^0 = X - m_x$, $Y^0 = Y - m_y$ и т. д.

Дисперсией скалярной случайной величины называется математическое ожидание квадрата модуля соответствующей центрированной случайной величины.

Дисперсию случайной величины X будем обозначать DX или, сокращенно, D_x *):

$$D_x = DX = M\,|X^0|^2 = M\,|X - m_x|^2. \qquad (14)$$

Средним квадратическим или *стандартным отклонением* σ_x случайной величины X называется положительный квадратный корень из ее дисперсии:

$$\sigma_x = \sqrt{D_x} = \sqrt{DX}. \qquad (15)$$

Чтобы получить формулу для вычисления дисперсии, достаточно положить в общей формуле (3) для математического ожидания функции случайной величины $\varphi(X) = |X^0|^2 = |X - m_x|^2$ и соответственно $\varphi(x) = |x - m_x|^2$. Если X — действительная случайная величина, то формула (3) даст

$$D_x = DX = \int_{-\infty}^{\infty} (x - m_x)^2 f(x)\,dx, \qquad (16)$$

где $f(x)$ — плотность величины X. Если X — комплексная случайная величина $X = X' + iX''$, то формула (3) даст

$$D_x = DX = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x' + ix'' - m_x|^2 f(x', x'')\,dx'\,dx'', \qquad (17)$$

где $f(x', x'')$ — совместная плотность случайных величин X' и X''.

---

*) Иногда, особенно в иностранной литературе, дисперсия величины X обозначается VX от английского variance (изменчивость).

Подставив в (16) выражение (2.24), получим формулу для дисперсии действительной дискретной случайной величины:

$$D_x = DX = \sum_{k=1}^{N} (x_k - m_x)^2 p_k. \qquad (18)$$

Подставив в (16) выражение (2.25), получим формулу для дисперсии непрерывно-дискретной случайной величины:

$$D_x = DX = \int_{-\infty}^{\infty} (x - m_x)^2 f_1(x)\,dx + \sum_{k=1}^{N} (x_k - m_x)^2 p_k. \qquad (19)$$

Пример 1. Найти математическое ожидание и дисперсию числа X появлений события A при n независимых опытах, если вероятность события A в каждом опыте равна p (биномиальное распределение — пример 2.1).

По формуле (2) находим

$$m_x = \sum_{m=0}^{n} m P_{m,n} = \sum_{m=0}^{n} m C_n^m p^m q^{n-m}, \qquad q = 1 - p.$$

Для вычисления этой суммы воспользуемся производящей функцией (1.25):

$$\varphi_n(u) = \sum_{m=0}^{n} P_{m,n} u^m = (q + pu)^n.$$

Дифференцируя эту формулу по u, получаем

$$\varphi_n'(u) = \sum_{m=0}^{n} m P_{m,n} u^{m-1} = np(q + pu)^{n-1}. \qquad (I)$$

Положив здесь $u = 1$ и учитывая, что $q + p = 1$, находим

$$\sum_{m=0}^{n} m P_{m,n} = \varphi_n'(1) = np.$$

Таким образом, $m_x = np$.

Для вычисления дисперсии воспользуемся формулой (17):

$$D_x = \sum_{m=0}^{n} (m - np)^2 P_{m,n} = \sum_{m=0}^{n} m^2 P_{m,n} - 2np \sum_{m=0}^{n} m P_{m,n} + n^2p^2 \sum_{m=0}^{n} P_{m,n}. \qquad (II)$$

Вторую сумму здесь мы уже вычислили, она равна np. Третья сумма равна 1 в силу основного свойства дискретных распределений (то же, конечно, получится, если положить в выражении производящей

функции (1.25) $u=1$). Для вычисления первой суммы умножим (I) на u и продифференцируем результат по u:

$$\frac{d}{du}\left[u\varphi_n'(u)\right]=\frac{d}{du}\sum_{m=0}^{n} mP_{m,n}u^m=\sum_{m=0}^{n} m^2P_{m,n}u^{m-1}=$$

$$=\frac{d}{du}\left[npu\,(q+pu)^{n-1}\right]=np\,(q+npu)\,(q+pu)^{n-2}.$$

Положив здесь $u=1$, находим

$$\sum_{m=0}^{n} m^2P_{m,n}=\frac{d}{du}\left[u\varphi_n'(u)\right]\Big|_{u=1}=np\,(q+np).$$

Подставив найденные выражения сумм в (II), получаем $D_x=npq$.

Пример 2. Найти математическое ожидание и дисперсию случайной величины X, распределенной по закону Пуассона (пример 2.3).

По формуле (2) находим математическое ожидание:

$$m_x=\sum_{m=0}^{\infty} mp_m=\sum_{m=1}^{\infty} m\frac{\mu^m}{m!}e^{-\mu}=\mu e^{-\mu}\sum_{m=1}^{\infty}\frac{\mu^{m-1}}{(m-1)!}=$$

$$=\mu e^{-\mu}\left(1+\frac{\mu}{1!}+\frac{\mu^2}{2!}+\ldots+\frac{\mu^k}{k!}+\ldots\right)=\mu e^{-\mu}e^{\mu}=\mu.$$

По формуле (18) находим дисперсию:

$$D_x=\sum_{m=0}^{\infty}(m-\mu)^2p_m=\sum_{m=2}^{\infty} m\,(m-1)\frac{\mu^m}{m!}e^{-\mu}-$$

$$-(2\mu-1)\sum_{m=1}^{\infty} m\frac{\mu^m}{m!}e^{-\mu}+\mu^2\sum_{m=0}^{\infty}\frac{\mu^m}{m!}e^{-\mu}. \qquad \text{(III)}$$

Вторую сумму мы уже вычислили, она равна μ. Третья сумма равна 1. Остается вычислить первую сумму:

$$\sum_{m=0}^{\infty} m\,(m-1)\frac{\mu^m}{m!}e^{-\mu}=\mu^2e^{-\mu}\sum_{m=2}^{\infty}\frac{\mu^{m-2}}{(m-2)!}=\mu^2.$$

Подставив найденные выражения сумм в (III), получаем $D_x=\mu$. Таким образом, математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, равны параметру μ в (1.46).

Пример 3. Для равномерно распределенной в интервале (a, b) случайной величины X (пример 2.5) формулы (3) и (17) дают

$$m_x=\int_a^b\frac{x\,dx}{b-a}=\frac{a+b}{2},\quad D_x=\int_a^b\left(x-\frac{a+b}{2}\right)^2\frac{dx}{b-a}=\frac{(b-a)^2}{12}.$$

Пример 4. Для случайной величины X, распределенной по показательному закону (пример 2.8), получаем

$$m_x = \int_0^{\infty} xke^{-kx}\,dx = \frac{1}{k}, \quad D_x = \int_0^{\infty} \left(x - \frac{1}{k}\right)^2 ke^{-kx}\,dx = \frac{1}{k^2}.$$

Пример 5. В условиях примера 2.14 математическое ожидание и дисперсия времени S работы системы после первого ремонта определяются по формулам (5) и (19):

$$m_s = \int_0^{t_0} s\lambda e^{-\lambda(t_0 - s)}ds + 0\cdot e^{-\lambda t_0} = t_0 - \frac{1}{\lambda}\left(1 - e^{-\lambda t_0}\right),$$

$$D_s = \int_0^{t_0} \left[s - t_0 + \frac{1}{\lambda}\left(1 - e^{-\lambda t_0}\right)\right]^2 \lambda e^{-\lambda(t_0 - s)}\,ds +$$

$$+\left[0 - t_0 + \frac{1}{\lambda}\left(1 - e^{-\lambda t_0}\right)\right]^2 e^{-\lambda t_0} = \frac{1}{\lambda^2}\left(1 - e^{-2\lambda t_0}\right) - \frac{2t_0}{\lambda}e^{-\lambda t_0}.$$

Пример 6. В качестве примера случайной величины, не имеющей математического ожидания и дисперсии, рассмотрим случайную величину X, распределенную по *закону Коши*, который определяется плотностью

$$f(x) = \frac{1}{\pi}\,\frac{b}{b^2 + (x-a)^2}. \tag{20}$$

В этом случае

$$m_x = \int_{-\infty}^{\infty} \frac{x}{\pi}\,\frac{b\,dx}{b^2 + (x-a)^2} = \frac{1}{\pi}\int_{-\infty}^{\infty} (a + bu)\frac{du}{1+u^2} = a + \frac{b}{\pi}\int_{-\infty}^{\infty} \frac{u\,du}{1+u^2}. \tag{IV}$$

Несобственный интеграл в этой формуле расходится, так так

$$\int_{-\infty}^{\infty} \frac{|u|\,du}{1+u^2} = 2\int_0^{\infty} \frac{u\,du}{1+u^2} = \infty.$$

Следовательно, случайная величина X не имеет математического ожидания. Однако если учесть симметрию распределения Коши относительно точки a и соответственно понимать интеграл в (IV) в смысле главного значения Коши [69, 109], то будем иметь

$$m_x = a + \frac{b}{\pi}\lim_{N\to\infty}\int_{-N}^{N} \frac{u\,du}{1+u^2} = a.$$

В этом случае дисперсия случайной величины X определится формулой

$$D_x = \frac{b}{\pi}\int_{-\infty}^{\infty} \frac{(x-a)^2\,dx}{b^2 + (x-a)^2} = \frac{b^2}{\pi}\int_{-\infty}^{\infty} \frac{u^2\,du}{1+u^2} = \infty.$$

Таким образом, случайная величина, распределенная по закону Коши, не имеет дисперсии даже в том случае, если ее математическое ожидание считать существующим и равным a из соображений симметрии.

2.2. Ковариация. *Ковариацией* скалярных случайных величин X и Y называется математическое ожидание произведения центрированной первой случайной величины и сопряженной центрированной второй случайной величины:

$$k_{xy} = MX^0\overline{Y^0}. \tag{21}$$

Чтобы получить формулу для вычисления ковариации действительных случайных величин X и Y, достаточно рассматривать $X^0Y^0 = (X - m_x)(Y - m_y)$ как функцию случайных величин X и Y и применить формулу (3):

$$k_{xy} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x - m_x)(y - m_y) f(x, y)\, dx\, dy, \tag{22}$$

где $f(x, y)$ — совместная плотность величин X и Y.

Чтобы получить формулу для вычисления ковариации комплексных случайных величин $X = X' + iX''$ и $Y = Y' + iY''$, достаточно рассматривать произведение $X^0\overline{Y^0}$ как функцию четырехмерного случайного вектора с координатами X', X'', Y', Y''. Тогда получим

$$k_{xy} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x' + ix'' - m_x)(y' - iy'' - m_y) \times$$

$$\times f(x', x'', y', y'')\, dx'\, dx''\, dy'\, dy'', \tag{23}$$

где $f(x', x'', y', y'')$ — совместная плотность случайных величин X', X'', Y', Y''.

За характеристику зависимости между двумя случайными величинами X и Y принимается отношение их ковариации к произведению их средних квадратических отклонений. Эта безразмерная величина называется *коэффициентом корреляции* величин X и Y:

$$r_{xy} = \frac{k_{xy}}{\sigma_x \sigma_y} = \frac{k_{xy}}{\sqrt{D_x D_y}}. \tag{24}$$

Таким образом, чтобы получить числовые характеристики двумерного случайного вектора, следует добавить к математическим ожиданиям и дисперсиям его координат еще их ковариацию или коэффициент корреляции.

Очевидно, что ковариация случайной величины X с самой собой равна ее дисперсии, $k_{xx}=D_x$, а ее коэффициент корреляции с самой собой равен единице, $r_{xx}=1$; $k_{xy}=D_x$, $r_{xy}=1$ и в более общем случае, когда $Y^0=X^0$, $m_y \neq m_x$.

Пример 7. Найти ковариацию координат случайного вектора, равномерно распределенного на прямоугольнике $|x|<a$, $|y|<b$ (пример 2.10).

Подставив в (22) выражение (2.20) плотности и имея в виду, что в данном случае $m_x=m_y=0$, получим

$$k_{xy}=\int_{-a}^{a}\int_{-b}^{b}\frac{xy}{4ab}\,dx\,dy=0.$$

Пример 8. Для двумерного случайного вектора с равномерным распределением внутри эллипса $x^2/a^2+y^2/b^2=1$ (пример 2.10) $m_x=m_y=0$ и

$$k_{xy}=\int_{-a}^{a}dx\int_{-\sqrt{1-x^2/a^2}}^{\sqrt{1-x^2/a^2}}\frac{xy\,dy}{\pi ab}=0.$$

Пример 9. Для двумерного случайного вектора с равномерным распределением внутри эллипса $x^2/a^2+(y-cx)^2/b^2=1$ $m_x=m_y=0$ и

$$k_{xy}=\int_{-a}^{a}dx\int_{cx-b\sqrt{1-x^2/a^2}}^{cx+b\sqrt{1-x^2/a^2}}\frac{xy}{\pi ab}\,dy=\frac{ca^2}{4}.$$

2.3. Коррелированные и некоррелированные случайные величины. Зависимость между случайными величинами, характеризуемая коэффициентом корреляции, называется *корреляцией*. Случайные величины называются *коррелированными*, если их коэффициент корреляции отличен от нуля. Случайные величины называются *некоррелированными*, если их коэффициент корреляции равен нулю.

Из (24) следует, что *случайные величины не коррелированы тогда и только тогда, когда их ковариация равна нулю*.

Легко видеть, что для некоррелированности случайных величин достаточно, чтобы их совместное распределение было симметрично относительно какой-нибудь прямой, параллельной одной из осей координат.

Пример 10. Ковариация случайных величин X и $Y = X^2$ в случае равномерного распределения X в интервале $(-a, a)$ равно 0:

$$k_{xy} = \int_{-a}^{a} \frac{x(x^2 - m_y)}{2a} dx = 0.$$

Следовательно, величины X и $Y = X^2$ не коррелированы, несмотря на то что они связаны самой жесткой зависимостью — функциональной.

2.4. Моменты первого и второго порядков. Математическое ожидание, дисперсия и ковариация представляют собой частные виды моментов случайных величин.

Моментом первого порядка (первым моментом) случайной величины называется ее математическое ожидание.

Моментом второго порядка (вторым моментом) скалярной (в общем случае комплексной) случайной величины X называется математическое ожидание квадрата ее модуля:

$$\gamma_x = M|X|^2. \tag{25}$$

Центральным моментом второго порядка величины X называется момент второго порядка центрированной величины $X^0 = X - m_x$, т. е. ее дисперсия.

Моментом второго порядка величины X *относительно точки* a называется момент второго порядка разности $X - a$:

$$\gamma_x(a) = M|X - a|^2. \tag{26}$$

Очевидно, что $\gamma_x = \gamma_x(0)$, $D_x = \gamma_x(m_x)$.

Смешанным моментом второго порядка скалярных случайных величин X и Y называется математическое ожидание произведения первой величины и сопряженной второй:

$$\gamma_{xy} = MX\bar{Y}. \tag{27}$$

Центральным смешанным моментом второго порядка величин X и Y называется смешанный второй момент центрированных случайных величин X^0 и Y^0, т. е. ковариация величин X и Y.

Смешанным моментом второго порядка величин X и Y *относительно точек* a *и* b называется смешанный второй момент разностей $X - a$ и $Y - b$:

$$\gamma_{xy}(a, b) = M(X - a)(\bar{Y} - \bar{b}). \tag{28}$$

Ясно, что $\gamma_{xy} = \gamma_{xy}(0, 0)$, $k_{xy} = \gamma_{xy}(m_x, m_y)$.

▶ Подставив в (25) и (27) $X = m_x + X^0$, $Y = m_y + Y^0$ и пользуясь свойствами (7) и (9) математических ожиданий, получим выражения моментов второго порядка через математические ожидания и центральные моменты второго порядка:

$$\gamma_{xx} = D_x + |m_x|^2, \tag{29}$$

$$\gamma_{xy} = k_{xy} + m_x \overline{m}_y. \tag{30}$$

Аналогично из (26) и (28) получаем

$$\gamma_{xx}(a) = D_x + |m_x - a|^2, \tag{31}$$

$$\gamma_{xy}(a, b) = k_{xy} + (m_x - a)(\overline{m}_y - \overline{b}). \quad ◀ \tag{32}$$

Таким образом, *все моменты второго порядка выражаются через математические ожидания случайных величин и их центральные моменты второго порядка.*

Формула (31) показывает, что *дисперсия случайной величины представляет собой наименьший из всех ее моментов второго порядка.*

§ 3. Моменты второго порядка случайных векторов

3.1. Момент второго порядка, ковариационная матрица, корреляционная матрица. *Моментом второго порядка (вторым моментом)* случайного вектора X называется матрица вторых моментов всех его координат:

$$\Gamma_x = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{bmatrix}, \qquad \gamma_{pq} = MX_p \overline{X}_q. \tag{33}$$

Момент второго порядка центрированного случайного вектора $X^0 = X - m_x$ называется *ковариационной матрицей* случайного вектора X:

$$K_x = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ k_{n1} & k_{n2} & \dots & k_{nn} \end{bmatrix}, \qquad k_{pq} = MX_p^0 \overline{X}_q^0. \tag{34}$$

Представив вектор X в форме матрицы-столбца, можем переписать определения второго момента и ковариационной матрицы случайного вектора X в виде

$$\Gamma_x = MXX^*, \qquad K_x = MX^0 X^{0*}, \tag{35}$$

где звездочка означает операцию транспонирования матрицы с заменой всех ее комплексных элементов соответствующими сопряженными числами.

Матрица, элементами которой служат коэффициенты корреляции координат $X_1, \ldots, X_n$ случайного вектора X, называется его *корреляционной матрицей*:

$$R_x = \begin{bmatrix} 1 & r_{12} & \ldots & r_{1n} \\ r_{21} & 1 & \ldots & r_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ r_{n1} & r_{n2} & \ldots & 1 \end{bmatrix}, \quad r_{pq} = \frac{k_{pq}}{\sqrt{k_{pp}k_{qq}}}. \tag{36}$$

▶ Подставив в первую формулу (35) $X = m_x + X^0$ и пользуясь свойствами (7) и (9) математических ожиданий, получаем соотношение между моментом второго порядка, ковариационной матрицей и математическим ожиданием случайного вектора:

$$\Gamma_x = K_x + m_x m_x^*. \blacktriangleleft \tag{37}$$

3.2. Взаимный момент второго порядка и взаимная ковариационная матрица. Обобщая определение (35), назовем *взаимным моментом второго порядка* (*вторым моментом*) двух случайных векторов X и Y матрицу (в общем случае прямоугольную)

$$\Gamma_{xy} = MXY^*. \tag{38}$$

Взаимной ковариационной матрицей или *ковариацией* случайных векторов X и Y назовем взаимный момент второго порядка соответствующих центрированных случайных векторов $X^0 = X - m_x$ и $Y^0 = Y - m_y$, $K_{xy} = MX^0Y^{0*}$.

Взаимные момент второго порядка, ковариационная матрица и математические ожидания векторов X и Y связаны соотношением, аналогичным (37):

$$\Gamma_{xy} = K_{xy} + m_x m_y^*. \tag{39}$$

Случайные векторы X и Y называются *коррелированными*, если $K_{xy} \neq 0$, и *некоррелированными*, если $K_{xy} = 0$. Из этого определения следует, что векторы X и Y не коррелированы тогда и только тогда, когда каждая координата одного из них не коррелирована со всеми координатами другого.

3.3. Операторы моментов второго порядка. Как известно, любая матрица определяет линейный оператор, отображающий одно пространство в другое (или в то же самое).

Линейный оператор Γ_x, определяемый матрицей Γ_x, называется *оператором момента второго порядка* случайного вектора X.

Ковариационным оператором случайного вектора X называется оператор момента второго порядка соответствующего центрированного случайного вектора X^0, т. е. оператор K_x, определяемый ковариационной матрицей K_x вектора X.

Аналогично определяются *взаимный оператор момента второго порядка* и *взаимный ковариационный оператор* двух случайных векторов X и Y.

В частном случае ковариационный оператор K_x скалярной случайной величины X представляет собой оператор умножения на дисперсию D_x этой случайной величины.

3.4. Свойства моментов второго порядка. 1) Из (33) — (35) следует, что *моменты второго порядка случайных векторов обладают эрмитовской симметрией* (*являются эрмитовыми*):

$$\Gamma_x^* = MX^{**}X^* = MXX^* = \Gamma_x, \qquad \gamma_{qp} = \bar{\gamma}_{pq}. \tag{40}$$

2) Из (35) следует, что *матрицы Γ_x и K_x неотрицательно определенны.*

▶ Действительно, для любого вектора (матрицы-столбца) u имеем

$$u^T\Gamma_x\bar{u} = M(u^TX)(X^*\bar{u}) = M(X^Tu)(\overline{X^Tu}) = M|X^Tu|^2 \geqslant 0, \tag{41}$$

причем знак равенства возможен тогда и только тогда, когда $X^Tu = 0$ с вероятностью 1, т. е. когда координаты случайного вектора X связаны линейной функциональной зависимостью. Действительно,

$$M|X^Tu|^2 = \int_{-\infty}^{\infty} |x^Tu|^2 f(x)\,dx = \int_{|x^Tu|>0} |x^Tu|^2 f(x)\,dx > 0,$$

если

$$P(|X^Tu| > 0) = \int_{|x^Tu|>0} f(x)\,dx > 0,$$

где неравенство под знаком интеграла указывает, что областью интегрирования служит часть пространства, в которой выполнено это неравенство. ◀

Таким образом, *момент второго порядка случайного вектора представляет собой эрмитову неотрицательно определенную матрицу.* В § 4 мы увидим, что всякая матрица, обладающая этими двумя свойствами, может служить ковариационной матрицей, а следовательно, и моментом второго порядка некоторого случайного вектора.

3) Легко убедиться также в том, что *при изменении порядка случайных векторов их взаимные второй момент и ковариационная матрица переходят в эрмитовски сопряжённые матрицы:*

$$\Gamma_{yx} = \Gamma_{xy}^*, \quad K_{yx} = K_{xy}^*. \tag{42}$$

4) Из неотрицательной определенности матрицы $\Gamma_x (K_x)$ следует, что *ее определитель действителен, неотрицателен и равен нулю тогда и только тогда, когда координаты случайного вектора X (соответственно $X^0 = X - m_x$) связаны линейными функциональными зависимостями.*

Действительно, определитель эрмитовой матрицы равен произведению ее собственных значений, а собственные значения неотрицательно определенной эрмитовой матрицы все неотрицательны. Отсюда следует, что определитель матрицы $\Gamma_x (K_x)$ неотрицателен и равен нулю тогда и только тогда, когда она имеет нулевое собственное значение.

Если нуль является m-кратным собственным значением матрицы $\Gamma_x (K_x)$ (т. е. m из n ее собственных значений равны нулю), то в пространстве значений вектора X существуют m взаимно ортогональных векторов $\varphi_1, \ldots, \varphi_m$, для которых

$$\Gamma_x \varphi_p = 0 \qquad (p = 1, \ldots, m).$$

Умножив это соотношение слева на φ_p^* и учитывая (35), получаем

$$\varphi_p^* \Gamma_x \varphi_x = M \varphi_p^* X X^* \varphi_p = M \mid \varphi_p^* X \mid^2 = 0.$$

Отсюда следует, что с вероятностью 1

$$\varphi_p^* X = 0 \qquad (p = 1, \ldots, m).$$

Таким образом, в случае m-кратного нулевого собственного значения матрицы Γ_x координаты вектора X связаны m независимыми линейными соотношениями. В этом случае распределение вектора X полностью сосредоточено на $(n - m)$-мерном подпространстве, образованном

$n-m$ собственными векторами матрицы Γ_x, соответствующими отличным от нуля собственным значениям (соответственно на $(n-m)$-мерном подпространстве, образованном $n-m$ собственными векторами матрицы K_x, сдвинутом на вектор m_x). Иными словами, *матрица* Γ_x (K_x) *имеет ранг* r *тогда и только тогда, когда распределение случайного вектора* X *полностью сосредоточено на* r-*мерном подпространстве (соответственно на* r-*мерном подпространстве, сдвинутом на вектор* m_x).

5) Выразив на основании (36) элементы ковариационной матрицы K_x формулой $k_{pq}=r_{pq}\sqrt{k_{pp}k_{qq}}$, убеждаемся в том, что и *определитель корреляционной матрицы неотрицателен и равен нулю тогда и только тогда, когда случайные величины* $X_1^0, \ldots, X_n^0$ *связаны линейными функциональными зависимостями*. При этом *ранг корреляционной матрицы равен* r *тогда и только тогда, когда распределение вектора* X *сосредоточено на* r-*мерном подпространстве, сдвинутом на вектор* m_x.

Таким образом, при любом n

$$\begin{vmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} \end{vmatrix} \geqslant 0, \qquad \begin{vmatrix} k_{11} & k_{12} & \ldots & k_{1n} \\ k_{21} & k_{22} & \ldots & k_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ k_{n1} & k_{n2} & \ldots & k_{nn} \end{vmatrix} \geqslant 0, \qquad (43)$$

и аналогичное неравенство справедливо для определителя корреляционной матрицы.

6) При $n=2$ неравенства (43) и аналогичное неравенство для корреляционной матрицы дают

$$|\gamma_{12}| \leqslant \sqrt{\gamma_{11}\gamma_{22}}, \quad |k_{12}| \leqslant \sqrt{k_{11}k_{22}}, \quad |r_{12}| \leqslant 1. \qquad (44)$$

Последнее неравенство убеждает нас в том, что *коэффициент корреляции случайных величин не может быть по модулю больше единицы, и его модуль равен единице тогда и только тогда, когда центрированные случайные величины связаны линейной зависимостью.*

7) Положив $X_1=X$, $X_2=Y$, перепишем первое неравенство (44) в виде

$$|MX\overline{Y}|^2 \leqslant M|X|^2 M|Y|^2. \qquad (45)$$

Знак равенства здесь имеет место тогда и только тогда, когда случайные величины X и Y связаны линейной зависимостью $c_1X+c_2Y=0$.

8) Заметим, что все установленные свойства матриц Γ_x и K_x присущи матрицам Γ_x и K_x, определяемым фор-

мулами (35), и в том случае, когда X представляет собой произвольную прямоугольную матрицу. В этом случае элементы матриц Γ_x и K_x определяются формулами

$$\gamma_{pq} = M \sum_{r=1}^{m} X_{pr} \overline{X}_{qr}, \quad k_{pq} = M \sum_{r=1}^{m} X_{pr}^0 \overline{X}_{qr}^0.$$

Вводя векторы (матрицы-столбцы)

$$X_p = [X_{p1} \ldots X_{pm}]^T \qquad (p = 1, \ldots, n),$$

можем представить формулы, определяющие γ_{pq} и k_{pq}, в виде

$$\gamma_{pq} = M X_p^T \overline{X}_q, \quad k_{pq} = M X_p^{0\,T} \overline{X}_q^0.$$

Тогда, положив $X_1 = X$, $X_2 = \overline{Y}$, перепишем первую формулу (44) в виде

$$| M X^T Y |^2 \leqslant M | X |^2 M | Y |^2, \tag{46}$$

где модули следует понимать как модули (евклидовы нормы) векторов:

$$| X |^2 = \sum_{p=1}^{m} | X_p |^2, \quad | Y |^2 = \sum_{p=1}^{m} | Y_p |^2. \tag{47}$$

Знак равенства в (46) имеет место тогда и только тогда, когда случайные векторы X и Y связаны линейной зависимостью $c_1 X + c_2 Y = 0$.

3.5. Линейные преобразования случайных векторов. Рассмотрим теперь случайный вектор Y, полученный линейным преобразованием вектора X:

$$Y = AX + a_0. \tag{48}$$

Его математическое ожидание определяется формулой (13), которую мы перепишем в виде

$$m_y = A m_x + a_0. \tag{49}$$

Вычитая (49) из (48), получаем $Y^0 = AX^0$. Пользуясь этим соотношением и определением (35), находим ковариационную матрицу вектора Y:

$$K_y = M Y^0 Y^{0*} = M A X^0 X^{0*} A^*.$$

Но в силу (11), (12) и (35)

$$M A X^0 X^{0*} A^* = A (M X^0 X^{0*}) A^* = A K_x A^*.$$

Следовательно,

$$K_y = AK_xA^*. \tag{50}$$

Эта формула показывает, что *при преобразовании случайного вектора X линейным оператором с матрицей A его ковариационная матрица умножается слева на матрицу A, а справа на эрмитовски сопряженную матрицу A^*.*

В частном случае, когда Y — скалярная случайная величина, A представляет собой матрицу-строку, $A = [a_1 \ \dots \ a_n]$, и формула (50) определяет дисперсию случайной величины Y:

$$D_y = \sum_{p,\,q=1}^{n} k_{pq} a_p \bar{a}_q. \tag{51}$$

В частном случае некоррелированных случайных величин $X_1, \dots, X_n$ эта формула принимает вид

$$D_y = \sum_{p=1}^{n} D_p |a_p|^2, \tag{52}$$

где $D_p = k_{pp}$ — дисперсия величины X_p $(p = 1, \dots, n)$. В частном случае при $n = 1$ отсюда следует, что *при умножении случайной величины на число ее дисперсия умножается на квадрат модуля этого числа.*

Аналогично выводится формула

$$K_{zu} = AK_{xy}B^* \tag{53}$$

для взаимной ковариационной матрицы случайных векторов

$$Z = AX + a_0, \qquad U = BY + b_0. \tag{54}$$

В частном случае при $Y = X$ формула (53) принимает вид

$$K_{zu} = AK_xB^*. \tag{55}$$

Эта формула определяет взаимную ковариационную матрицу случайных векторов Z и U, полученных двумя разными линейными преобразованиями одного и того же случайного вектора X:

$$Z = AX + a_0, \qquad U = BX + b_0. \tag{56}$$

В частном случае скалярных величин Z и U матрицы A и B представляют собой матрицы-строки, $A = [a_1 \ \dots \ a_n]$, $B = [b_1 \ \dots \ b_n]$, и формула (55) определяет

ковариацию случайных величин Z и U:

$$k_{zu} = \sum_{p,\, q=1}^{n} k_{pq} a_p \bar{b}_q. \qquad (57)$$

В частном случае некоррелированных координат $X_1, \ldots$ $\ldots, X_n$ вектора X эта формула принимает вид

$$k_{zu} = \sum_{p=1}^{n} D_p a_p \bar{b}_p, \qquad (58)$$

где $D_p = k_{pp} (p = 1, \ldots, n)$.

Чтобы привести формулу (50) в случае некоррелированных величин $X_1, \ldots, X_n$ к удобному для приложений виду, будем рассматривать p-й столбец матрицы A как матрицу-столбец a_p:

$$a_p = [a_{1p}\ a_{2p} \ldots a_{mp}]^T \qquad (p = 1, \ldots, n).$$

Тогда формула (48), определяющая вектор Y, перепишется в виде

$$Y = \sum_{p=1}^{n} X_p a_p + a_0. \qquad (59)$$

Формула (50), определяющая ковариационную матрицу вектора Y, в силу того, что $k_{pq} = 0$ при $q \neq p$, примет вид

$$K_y = \sum_{p=1}^{n} D_p a_p a_p^*, \qquad (60)$$

где $D_p = k_{pp}$—дисперсия величины X_p, т. е. p-й элемент диагональной матрицы $K_x (p = 1, \ldots, n)$.

Формула (60), конечно, может быть выведена непосредственно из (59) с учетом свойств (7), (10), (11) и (12) математических ожиданий.

К аналогичной форме приводится в случае вектора X с некоррелированными координатами и формула (55).

Заметим, что в случае линейных преобразований без сдвига (48), (54), (56), когда $a_0 = b_0 = 0$, соотношения (50), (53), (55) справедливы также и для вторых моментов Г. Следует, однако, помнить, что если a_0, $b_0 \neq 0$, то формулы (50), (53), и (55) справедливы только для ковариационных матриц K.

Пример 11. Найти математическое ожидание и дисперсию частоты $\hat{P}$ события A при n независимых опытах, если вероятность A в каждом опыте равна p.

Пусть X_k — число появлений события A при одном k-м опыте. Эта величина имеет два возможных значения 0 и 1, вероятности которых равны соответственно $q = 1 - p$ и p. Поэтому ее математическое ожидание m_k и дисперсия D_k определяются по формулам (2) и (18):

$$m_k = 0 \cdot q + 1 \cdot p = p,$$
$$D_k = (0 - p)^2 \cdot q + (1 - p)^2 \cdot p = p^2 \cdot q + q^2 \cdot p = pq.$$

Таким образом, все величины $X_1, \ldots, X_n$ имеют одинаковые математические ожидания и дисперсии, равные соответственно p и pq. Частота события при n опытах определяется формулой

$$\hat{P} = \frac{1}{n} \sum_{k=1}^{n} X_k,$$

так как число его появлений при n опытах равно сумме чисел его появлений в отдельных опытах, т. е. $X_1 + \ldots + X_n$. Таким образом, частота $\hat{P}$ представляет собой линейную функцию случайных величин $X_1, \ldots, X_n$. Поэтому для нахождения ее математического ожидания можно применить формулу (49):

$$M\hat{P} = \frac{1}{n} \sum_{k=1}^{n} MX_k = \frac{1}{n}(np) = p.$$

Для вычисления дисперсии частоты $\hat{P}$ заметим, что на основании теоремы, которая будет доказана в п. 4.2.4, из независимости опытов вытекает некоррелированность случайных величин $X_1, \ldots, X_n$. Поэтому для вычисления дисперсии можно применить формулу (52):

$$D\hat{P} = \frac{1}{n^2} \sum_{k=1}^{n} DX_k = \frac{1}{n^2}(npq) = \frac{pq}{n}.$$

Этот же результат получится, если воспользоваться результатами вычислений математического ожидания и дисперсии числа X появлений события A при n опытах в примере 1 и заметить, что в силу того, что $\hat{P} = X/n$,

$$M\hat{P} = \frac{1}{n} MX, \qquad D\hat{P} = \frac{1}{n^2} DX.$$

Наконец, заметив, что $X = X_1 + \ldots + X_n$, можно значительно проще получить результаты примера 1 по формулам для математического ожидания и дисперсии суммы некоррелированных случайных величин.

П р и м е р 12. Найти математическое ожидание и дисперсию среднего арифметического значения $\overline{X}$ действительных случайных величин $X_1, \ldots, X_n$.

По формулам (49) и (51) находим

$$M\overline{X} = \frac{1}{n} \sum_{p=1}^{n} MX_p, \quad D\overline{X} = \frac{1}{n^2} \sum_{p,\ q=1}^{n} k_{pq},$$

где $k_{pq}=MX_p^0X_q^0$. В частном случае некоррелированных величин $X_1, \ldots, X_n$ последняя формула принимает вид

$$D\overline{X}=\frac{1}{n^2}\sum_{p=1}^{n} D_p,$$

где $D_p=k_{pp}$ $(p=1, \ldots, n)$. Наконец, в случае, когда величины $X_1, \ldots, X_n$ не коррелированы и имеют одинаковые математические ожидания и дисперсии, равные соответственно m_x и D_x, полученные формулы дают

$$M\overline{X}=\frac{1}{n}(nm_x)=m_x, \qquad D\overline{X}=\frac{1}{n^2}(nD_x)=\frac{D_x}{n}.$$

Пример 13. Найти математическое ожидание и ковариационную матрицу случайного вектора Z, координаты которого $Z_1, \ldots, Z_n$ представляют собой числа появлений значений $x_1, \ldots, x_N$ дискретной случайной величины X при n независимых опытах, если в каждом опыте $P(X=x_\nu)=p_\nu$ $(\nu=1, \ldots, N)$.

Пусть $Z^{(i)}$ — случайный вектор, координатами которого служат числа появлений значений $x_1, \ldots, x_N$ в одном i-м опыте. Очевидно, что возможными значениями вектора $Z^{(i)}$ служат единичные векторы N-мерного пространства

$$e_1=\{1, 0, \ldots, 0\}, \; e_2=\{0, 1, \ldots, 0\}, \; \ldots, \; e_N=\{0, \ldots, 1\},$$

так как в одном опыте появляется только одно из значений $x_1, \ldots, x_N$. Вероятности этих значений равны $p_1, \ldots, p_N$. Следовательно, математическое ожидание вектора $Z^{(i)}$ равно

$$MZ^{(i)}=\sum_{\nu=1}^{N} e_\nu p_\nu=\{p_1, \ldots, p_N\}=p.$$

Таким образом, математическое ожидание вектора $Z^{(i)}$ представляет собой вектор p, координатами которого служат вероятности $p_1, \ldots$ $\ldots, p_N$ значений $x_1, \ldots, x_N$ случайной величины X. Для определения ковариационной матрицы вектора $Z^{(i)}$ найдем сначала его момент второго порядка Γ'. Заметив, что ν-я координата вектора $Z^{(i)}$ равна 1 с вероятностью p_ν и 0 с вероятностью $1-p_\nu$, находим диагональные элементы матрицы Γ':

$$\gamma'_{\nu\nu}=1^2\cdot p_\nu+0^2\cdot(1-p_\nu)=p_\nu.$$

Так как никакие две координаты вектора $Z^{(i)}$ не могут быть одновременно отличными от 0, то их произведение равно 0 с вероятностью 1. Поэтому $\gamma'_{\nu\mu}=0$ при $\mu\neq\nu$. Определив моменты второго порядка координат вектора $Z^{(i)}$, можно найти их дисперсии и ковариации по формулам (29) и (30):

$$k'_{\nu\nu}=p_\nu-p_\nu^2=p_\nu(1-p_\nu),$$
$$k'_{\nu\mu}=-p_\nu p_\mu, \qquad \mu\neq\nu.$$

Таким образом, ковариационная матрица K вектора $Z^{(i)}$ определяется формулой

$$K=\Lambda-pp^T,$$

где Λ — диагональная матрица с элементами $p_1, \ldots, p_N$, а p — вектор-столбец с теми же элементами.

Заметив теперь, что $Z=Z^{(1)}+\ldots+Z^{(n)}$, находим математическое ожидание и ковариационную матрицу вектора Z:

$$m_z=np, \qquad K_z=n(\Lambda-pp^T).$$

§ 4. Канонические разложения случайных векторов

4.1. Разложение по собственным векторам. Во многих задачах практики удобно оперировать случайными векторами с некоррелированными координатами. В связи с этим возникает задача приведения любого случайного вектора к вектору с некоррелированными координатами (*декорреляции* координат случайного вектора). Эта задача равноценна задаче приведения ковариационной матрицы к диагональной форме.

▶ Из линейной алгебры известно, что любая эрмитовская (симметричная) матрица может быть приведена к диагональной форме унитарным (ортогональным) преобразованием. Пусть $\lambda_1, \ldots, \lambda_n$ — собственные значения матрицы K_x (неотрицательные в силу неотрицательной определенности K_x), $\varphi_1, \ldots, \varphi_n$ — соответствующие ортонормированные собственные векторы. Тогда будут справедливы соотношения

$$K_x\varphi_p=\lambda_p\varphi_p \qquad (p=1, \ldots, n), \tag{61}$$

$$(\varphi_p, \varphi_q)=\varphi_q^*\varphi_p=\delta_{pq} \qquad (p, q=1, \ldots, n)\text{*)}. \tag{62}$$

Пусть Φ — квадратная матрица, столбцами которой служат собственные векторы:

$$\Phi=[\varphi_1\ldots\varphi_n]. \tag{63}$$

Покажем, что случайный вектор

$$U=\Phi^*X^0 \tag{64}$$

имеет некоррелированные координаты. Для этого найдем по формуле (50) его ковариационную матрицу:

$$K_u=\Phi^*K_x\Phi=\Phi^*K_x[\varphi_1\ldots\varphi_n]=\Phi^*[K_x\varphi_1\ldots K_x\varphi_n],$$

---

*) В первой части формулы скалярное произведение записано в обычной форме, а во второй — в матричной, когда векторы представлены в форме матриц-столбцов; $\delta_{pp}=1$, $\delta_{pq}=0$ при $p\neq q$ (символ Кронекера).

или, в силу (61),

$$K_u = \begin{bmatrix} \varphi_1^* \\ \cdot \\ \cdot \\ \cdot \\ \varphi_n^* \end{bmatrix} [\lambda_1\varphi_1 \ldots \lambda_n\varphi_n] = \begin{bmatrix} \lambda_1\varphi_1^*\varphi_1 \ldots \lambda_n\varphi_1^*\varphi_n \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \lambda_1\varphi_n^*\varphi_1 \ldots \lambda_n\varphi_n^*\varphi_n \end{bmatrix}.$$

Отсюда, принимая во внимание (62), получаем

$$K_u = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}.$$

Таким образом, ковариационная матрица случайного вектора U диагональна. Следовательно, его координаты не коррелированы и их дисперсии равны соответствующим собственным значениям ковариационной матрицы K_x, $DU_p = \lambda_p \, (p = 1, \ldots, n)$.

Из (64) вытекает следующая формула для координат вектора U:

$$U_p = \varphi_p^* X^0 = (X^0, \varphi_p) \qquad (p = 1, \ldots, n). \tag{65}$$

Эта формула показывает, что координаты вектора U представляют собой проекции случайного вектора X^0 на собственные векторы.

Наконец, представив вектор $X^0 = X - m_x$ разложением по базису $\{\varphi_1, \ldots, \varphi_n\}$, получаем

$$X = m_x + \sum_{p=1}^{n} (X^0, \varphi_p)\varphi_p = m_x + \sum_{p=1}^{n} U_p\varphi_p. \tag{66}$$

Эта формула дает разложение случайного вектора X по собственным векторам его ковариационной матрицы. Сравнив (66) с (59) и применив формулу (60), получим разложение ковариационной матрицы случайного вектора X по собственным векторам:

$$K_x = \sum_{p=1}^{n} \lambda_p \varphi_p \varphi_p^*. \blacktriangleleft \tag{67}$$

Случайные величины $U_1\varphi_1, \ldots, U_n\varphi_n$ в разложении (66) по собственным векторам называются *главными компонентами* случайного вектора X. Это название объясняется тем, что среди проекций вектора X на все возможные направления максимальную дисперсию, равную наибольшему собственному значению λ_1, имеет про-

екция на собственный вектор φ_1 и при любом s среди проекций вектора $X^0-U_1\varphi_1-\ldots-U_{s-1}\varphi_{s-1}$ на все возможные направления максимальной дисперсией, равной s-му но величине собственному значению λ_s, обладает проекция на собственный вектор φ_s (экстремальное свойство собственных значений) [96, т. 3, ч. 1]:

$$\lambda_1=\sup_{|\varphi|=1}\varphi^*K_x\varphi,$$

$$\lambda_s=\sup_{|\varphi|=1}\varphi^*\left(K_x-\sum_{p=1}^{s-1}\lambda_p\varphi_p\varphi_p^*\right)\varphi.$$

4.2. Вычисление собственных значений и собственных векторов. Для получения разложения случайного вектора по собственным векторам необходимо найти собственные значения и собственные векторы его ковариационной матрицы. Эта задача сводится к решению алгебраического уравнения n-й степени

$$|K_x-\lambda I|=0\text{ *)}$$

и соответствующих систем однородных линейных уравнений (61), что связано со значительными вычислительными трудностями. Поэтому для вычисления собственных значений и собственных векторов обычно применяют приближенные вычислительные методы линейной алгебры [107]. Одним из наиболее удобных для вычисления на ЭВМ является *метод последовательных поворотов осей координат*, приводящих наибольшие по модулю внедиагональные элементы ковариационной матрицы к нулю [59] **).

Пусть k_{pq}, $p<q$, — наибольший по модулю элемент ковариационной матрицы K_x действительного случайного вектора $X=[X_1\ \ldots\ X_n]^T$. Повернем оси координат в плоскости, образованной единичными векторами (ортами) e_p и e_q, так, чтобы новые единичные векторы e'_p и e'_q были направлены по осям проекции эллипсоида $u^TK_xu=\text{const}$ на эту плоскость, т. е. чтобы проекции X'_p и X'_q вектора X на e'_p и e'_q были некоррелированными. Обозначив угол поворота θ, будем иметь

$$X'_p=X_p\cos\theta-X_q\sin\theta,\qquad X'_q=X_p\sin\theta+X_q\cos\theta.$$

---

*) Через I мы обозначаем единичную матрицу.

**) В [94] для вычисления собственных значений и собственных векторов этим методом служит программа EIGEN.

При этом элементы k_{pq}, k_{pp}, k_{qq}, k_{pr}, k_{qr} $(r \neq p,\ q)$ преобразуются по формулам

$$k'_{pq} = k_{pq} \cos 2\theta + \frac{1}{2}(k_{pp} - k_{qq}) \sin 2\theta,$$

$$k'_{pp} = \frac{1}{2}(k_{pp} + k_{qq}) + \frac{1}{2}(k_{pp} - k_{qq}) \cos 2\theta - k_{pq} \sin 2\theta,$$

$$k'_{qq} = \frac{1}{2}(k_{pp} + k_{qq}) - \frac{1}{2}(k_{pp} - k_{qq}) \cos 2\theta + k_{pq} \sin 2\theta,$$

$$\left.\begin{aligned} k'_{pr} &= k_{pr} \cos \theta - k_{qr} \sin \theta, \\ k'_{qr} &= k_{pr} \sin \theta + k_{qr} \cos \theta, \end{aligned}\right\} \quad (r \neq p,\ q),$$

а остальные элементы остаются неизменными. Единичными векторами e'_p и e'_q новой системы координат, полученной в результате поворота, служат векторы

$$e'_p = e_p \cos \theta - e_q \sin \theta, \qquad e'_q = e_p \sin \theta + e_q \cos \theta.$$

Определив θ из условия $k'_{pq} = 0$, получим

$$\operatorname{tg} 2\theta = -\frac{2k_{pq}}{k_{pp} - k_{qq}}.$$

При этом сумма квадратов внедиагональных элементов матрицы K_x

$$\nu^2 = 2 \sum_{s=2}^{n} \sum_{r=1}^{s-1} k_{rs}^2$$

уменьшится на величину k_{pq}^2, так как $k'^2_{pr} + k'^2_{qr} = k_{pr}^2 + k_{qr}^2$ при любом $r \neq p, q$. Отсюда следует, что при повторении подобных поворотов, аннулирующих каждый раз наибольший по модулю внедиагональный элемент матрицы K_x, сумма квадратов внедиагональных элементов убывает и имеет пределом 0. Это доказывает, что матрица K_x в результате неограниченной последовательности таких поворотов приводится к диагональной форме, а матрица, столбцами которой служат каждый раз орты новой системы координат, стремится к матрице Φ, столбцы которой представляют собой собственные векторы матрицы K_x.

Таким образом, метод последовательных поворотов дает возможность определить собственные значения и собственные векторы с любой степенью точности.

П р и м е р 14. Разложить по собственным векторам случайный вектор X с математическим ожиданием $m_x = 0$ и ковариационной матрицей

$$K_x = \begin{bmatrix} 1{,}94 & 1{,}80 & 1{,}58 & 1{,}11 & 0{,}61 \\ 1{,}80 & 1{,}90 & 1{,}63 & 0{,}87 & 0{,}48 \\ 1{,}58 & 1{,}63 & 1{,}74 & 0{,}94 & 0{,}59 \\ 1{,}11 & 0{,}87 & 0{,}94 & 1{,}30 & 0{,}36 \\ 0{,}61 & 0{,}48 & 0{,}59 & 0{,}36 & 1{,}10 \end{bmatrix}.$$

Применив метод последовательных поворотов, по стандартной подпрограмме EIGEN на ЭВМ, приводим матрицу K_x к виду *)

$$K'_x = \begin{bmatrix} 6{,}05004 & 0{,}0^{12}914 & 0{,}0^{10}1523 & 0{,}0^{8}246 & 0{,}0^{6}294 \\ 0{,}0^{12}914 & 0{,}900250 & -0{,}0^{6}422 & -0{,}0^{12}910 & 0{,}0^{10}1331 \\ 0{,}0^{10}1523 & -0{,}0^{6}422 & 0{,}713409 & -0{,}0^{10}575 & -0{,}0^{9}345 \\ 0{,}0^{8}246 & -0{,}0^{12}910 & -0{,}0^{10}575 & 0{,}260251 & 0{,}0^{13}568 \\ 0{,}0^{6}294 & 0{,}0^{10}1331 & -0{,}0^{9}345 & 0{,}0^{13}568 & 0{,}0560337 \end{bmatrix}.$$

При этом получена следующая матрица единичных векторов преобразованной системы координат:

$$\Phi = \begin{bmatrix} 0{,}550070 & -0{,}0903568 & -0{,}0148574 & -0{,}556621 & -0{,}615800 \\ 0{,}532374 & -0{,}247680 & -0{,}360632 & -0{,}195935 & 0{,}697700 \\ 0{,}505474 & -0{,}0373498 & -0{,}174132 & 0{,}802011 & -0{,}263734 \\ 0{,}341577 & 0{,}00976712 & 0{,}909557 & 0{,}0580490 & 0{,}229269 \\ 0{,}204499 & 0{,}963846 & -0{,}110029 & -0{,}0720412 & 0{,}109015 \end{bmatrix}.$$

Таким образом, собственные значения ковариационной матрицы K_x можно считать равными диагональным элементам матрицы K'_x. Ограничиваясь четырьмя знаками после запятой, получим $\lambda_1 = 6{,}0500$, $\lambda_2 = 0{,}9002$, $\lambda_3 = 0{,}7134$, $\lambda_4 = 0{,}2603$, $\lambda_5 = 0{,}0560$, а за собственные векторы можно принять соответствующие столбцы матрицы Ф. Ограничиваясь при определении координат собственных векторов также четырьмя знаками после запятой, получим разложение случайного вектора X по собственным векторам в виде

$$\begin{aligned} X = {} & U_1 [\;0{,}5501 \quad 0{,}5324 \quad 0{,}5055 \quad 0{,}3416 \quad 0{,}2045]^T + \\ & + U_2 [-0{,}0904 \quad -0{,}2477 \quad -0{,}0373 \quad 0{,}0098 \quad 0{,}9638]^T + \\ & + U_3 [-0{,}0149 \quad -0{,}3606 \quad -0{,}1741 \quad 0{,}9096 \quad -0{,}1100]^T + \\ & + U_4 [-0{,}5566 \quad -0{,}1959 \quad 0{,}8020 \quad 0{,}0580 \quad -0{,}0720]^T + \\ & + U_5 [-0{,}6158 \quad 0{,}6977 \quad -0{,}2637 \quad 0{,}2293 \quad 0{,}1090]^T. \end{aligned}$$

Контрольное вычисление матрицы K_x по формуле (67) дает расхождение с исходной матрицей K_x, не превышающее двух единиц четвертого знака после запятой. Относительные ошибки при этом не превосходят 0,04%. Это дает основание считать, что собственные значения и собственные векторы матрицы K_x найдены с достаточной степенью точности.

4.3. Каноническое разложение. Мы видим, что определение собственных значений и собственных векторов всегда связано с громоздкими вычислениями. Между тем ковариационная матрица может быть приведена к диагональной форме бесчисленным множеством других способов, если отказаться от требования унитарности преобразования.

▶ Пусть $f_1, \ldots, f_n$ — произвольные линейно независимые векторы. Определим случайные величины $V_p = f_p^T X^0$

*) Показатель степени у нуля показывает число нулей перед первой значащей цифрой. Например, $0{,}0^{6}294 = 0{,}000000294$.

($p=1, \ldots, n$). На основании (50) и (55) дисперсии и ковариации величин $V_1, \ldots, V_n$ определяются формулой

$$MV_p\bar{V}_q = f_p^T K_x \bar{f}_q \qquad (p, q = 1, \ldots, n).$$

Отсюда видно, что величины $V_1, \ldots, V_n$ не коррелированы тогда и только тогда, когда $f_p^T K_x \bar{f}_q = 0$ при $q \neq p$.

Чтобы определить векторы $f_1, \ldots, f_n$, удовлетворяющие условию $f_p^T K_x \bar{f}_q = 0$, $q \neq p$, возьмем произвольные линейно независимые векторы $g_1, \ldots, g_n$ и положим

$$f_1 = g_1, \quad D_1 = DV_1 = f_1^T K_x \bar{f}_1, \quad x_1 = D_1^{-1} K_x \bar{f}_1.$$

Тогда будем иметь $f_1^T x_1 = D_1^{-1} f_1^T K_x \bar{f}_1 = 1$. Положим теперь

$$f_2 = g_2 + c_{21} f_1$$

и определим c_{21} из условия $f_2^T K_x \bar{f}_1 = 0$ или, что то же, $f_2^T x_1 = 0$. Так как

$$f_2^T x_1 = g_2^T x_1 + c_{21} f_1^T x_1 = g_2^T x_1 + c_{21},$$

то условие $f_2^T x_1 = 0$ дает $c_{21} = -g_2^T x_1$. Определив таким образом вектор

$$f_2 = g_2 - (g_2^T x_1) f_1,$$

дисперсию $D_2 = f_2^T K_x \bar{f}_2$ величины $V_2 = f_2^T X^0$ и вектор $x_2 = D_2^{-1} K_x \bar{f}_2$, будем иметь

$$f_2^T x_2 = D_2^{-1} f_2^T K_x \bar{f}_2 = 1$$

и в силу эрмитовой симметрии ковариационной матрицы

$$f_1^T x_2 = D_2^{-1} f_1^T K_x \bar{f}_2 = D_2^{-1} \overline{f_2^T K_x \bar{f}_1} = 0.$$

Продолжая этот процесс, определим последовательно векторы

$$f_1 = g_1, \; f_s = g_s - (g_s^T x_1) f_1 - \ldots - (g_s^T x_{s-1}) f_{s-1}, \tag{68}$$

случайные величины

$$V_s = f_s^T X^0, \tag{69}$$

их дисперсии

$$D_s = f_s^T K_x \bar{f}_s \tag{70}$$

и векторы

$$x_s = \frac{1}{D_s} K_x \bar{f}_s. \tag{71}$$

При этом векторы f_s и x_s будут удовлетворять условиям $f_p^T x_q = x_q^T f_p = \delta_{pq}$ (т. е. $\delta_{pq} = 0$ при $q \neq p$ и $\delta_{pq} = 1$ при

$q=p$), в силу которых величины V_p и V_q будут некоррелированными при $q \neq p$.

На основании (68) формулы (69), (70) и (71) можно переписать в виде

$$V_s = g_s^T \left(X^0 - \sum_{p=1}^{s-1} V_p x_p \right), \quad D_s = g_s^T K_x^{(s)} \bar{g}_s,$$

$$x_s = \frac{1}{D_s} K_x^{(s)} \bar{g}_s, \qquad (72)$$

где

$$K_x^{(s)} = K_x - \sum_{p=1}^{s-1} D_p x_p x_p^* \qquad (73)$$

— ковариационная матрица случайного вектора

$$X_s = X^0 - \sum_{p=1}^{s-1} V_p x_p.$$

Формулы (72) и (73) дают возможность последовательно отщеплять от вектора X^0 некоррелированные слагаемые $V_s x_s$, каждое из которых не коррелировано с соответствующим остатком X_{s+1} (и всеми последующими остатками), не вычисляя векторов f_s.

Если ранг матрицы K_x равен размерности пространства n, то мы получим таким путем n векторов $f_1, \ldots, f_n$ и ни одна из величин $D_1, \ldots, D_n$ не будет равной 0. При этом формула (71) определит n векторов $x_1, \ldots, x_n$.

Если ранг r матрицы K_x меньше размерности пространства n, то при некотором $p \leqslant r+1$ получится $K_x \bar{f}_p = K_x^{(p)} \bar{g}_p = 0$, $D_p = 0$, и вектор x_p останется неопределенным. В таком случае можно продолжить вычисления по формулам (68)—(72), приняв $x_p = 0$. Поступая так, мы найдем в результате n векторов $f_1, \ldots, f_n$ и r отличных от нуля векторов x_s. Изменив в случае необходимости нумерацию векторов f_s и x_s, можем считать отличными от нуля векторы $x_1, \ldots, x_r$. Тогда векторам $f_1, \ldots, f_r$ будут соответствовать некоррелированные случайные величины $V_1 = f_1^T X^0, \ldots, V_r = f_r^T X^0$ с отличными от нуля дисперсиями $D_1, \ldots, D_r$, а дисперсии величин $V_{r+1} = f_{r+1}^T X^0, \ldots, V_n = f_n^T X^0$ будут равны 0. Вследствие этого величины $V_{r+1}, \ldots, V_n$ равны 0 с вероятностью 1 и координаты случайного вектора X^0 связаны $n-r$ линейными зависимостями $f_{r+1}^T X^0 = \ldots = f_n^T X^0 = 0$ в полном соответствии с результатом § 3.

Рассмотрим теперь случайный вектор

$$Y = X^0 - \sum_{p=1}^{r} V_p x_p.$$

Так как $f_s^T X^0 = V_s$, $f_s^T x_p = \delta_{sp}$ $(s = 1, \ldots, n;\ p = 1, \ldots, r)$ и $V_{r+1} = \ldots = V_n = 0$ почти наверное при $r < n$, то

$$f_s^T Y = f_s^T X^0 - V_s = 0 \qquad (s = 1, \ldots, r),$$
$$f_s^T Y = f_s^T X^0 = V_s = 0 \text{ п. н.} \quad (s = r+1, \ldots, n).$$

Таким образом, координаты вектора Y связаны n линейными зависимостями. Вводя матрицу F, столбцами которой служат векторы f_s,

$$F = [f_1 \ldots f_n],$$

можем написать эти зависимости в виде

$$F^T Y = 0.$$

Очевидно, что векторы $f_1, \ldots, f_n$ линейно независимы в силу линейной независимости исходных векторов $g_1, \ldots, g_n$. Поэтому определитель матрицы F не может быть равным 0. Следовательно, $Y = 0$ (почти наверное в случае, когда $r < n$). Таким образом,

$$X = m_x + \sum_{p=1}^{r} V_p x_p. \tag{74}$$

Это равенство точное при $r = n$ и выполняется с вероятностью 1 при $r < n$.

Таким образом, мы выразили случайный вектор X через случайный вектор V с некоррелированными координатами $V_1, \ldots, V_r$ (и, следовательно, с диагональной ковариационной матрицей). Сравнив (74) с (59) и применив формулу (60), получаем соответствующее выражение для ковариационной матрицы случайного вектора X:

$$K_x = \sum_{p=1}^{n} D_p x_p x_p^*. \blacktriangleleft \tag{75}$$

Представление случайного вектора X в виде линейной функции некоррелированных случайных величин (74) называется его *каноническим разложением*. Соответствующее разложение (75) называется каноническим разложением ковариационной матрицы случайного вектора X. Векторы $x_1, \ldots, x_r$ называются *координатными векторами* канонического разложения.

4.4. Различные формы канонического разложения. Из неоднозначности решения задачи приведения ковариационной матрицы к диагональной форме (произвольности выбора исходных линейно независимых векторов $g_1, \ldots, g_n$) следует, что *для любого случайного вектора с конечным моментом второго порядка существует бесчисленное множество канонических разложений.*

Разложение по собственным векторам является одним из канонических разложений. Оно получается, если за исходные векторы $g_1, \ldots, g_n$ взять сопряженные собственные векторы $\bar{\varphi}_1, \ldots, \bar{\varphi}_n$. Тогда будем иметь в силу (61) и (62)

$$f_p = \bar{\varphi}_p, \quad V_p = \varphi_p^* X^0 = U_p, \quad D_p = \lambda_p, \quad x_p = \varphi_p \quad (p = 1, \ldots, n)$$

и формулы (74) и (75) примут соответственно вид (66) и (67) (если $r < n$, то $\lambda_{r+1} = \ldots = \lambda_n = 0$ и соответственно $U_{r+1} = \ldots = U_n = 0$ п. н.).

Очевидно, что, вводя соответствующие множители в выражения случайных величин V_p и обратные множители в выражения векторов x_p, можно сделать дисперсии величин $V_1, \ldots, V_r$ равными любым заданным числам, например сделать все их равными 1, или добиться того, чтобы все векторы $x_1, \ldots, x_r$ были единичными. В последнем случае сумма дисперсий величин $V_1, \ldots, V_r$ будет равна сумме дисперсий координат $X_1, \ldots, X_n$ случайного вектора X, т. е. следу его ковариационной матрицы:

$$\sum_{p=1}^{r} D_p = \sum_{h=1}^{n} DX_h = \sum_{h=1}^{n} k_{hh} = \operatorname{tr} K_x.$$

Действительно, из (75) в этом случае следует, что

$$\operatorname{tr} K_x = \sum_{h=1}^{n} k_{hh} = \sum_{h=1}^{n} \sum_{p=1}^{r} D_p \, |x_{ph}|^2 = \sum_{p=1}^{r} D_p \sum_{h=1}^{n} |x_{ph}|^2 = \sum_{p=1}^{r} D_p.$$

В частности, для разложения по собственным векторам всегда

$$\sum_{p=1}^{n} \lambda_p = \sum_{h=1}^{n} k_{hh} = \operatorname{tr} K_x$$

(это следует также из известного факта, что при ортогональном преобразовании след матрицы не изменяется).

4.5. Простейший способ получения канонического разложения. Вычисления по формулам (68)—(72) будут проще, если принять за исходные векторы $g_1, \ldots, g_n$ единичные векторы осей координат $e_1, \ldots, e_n$. В этом случае

$$f_1^T = [1\ 0 \ldots 0],\ D_1 = k_{11},\ x_{1q} = D_1^{-1} k_{q1}\ (q = 1, \ldots, n)$$

и формулы (68)—(72) последовательно дают

$$\begin{aligned}
&D_s = k_{ss} - \sum_{q=1}^{s-1} D_q |x_{qs}|^2,\\
&x_{sp} = 0 \quad (p = 1, \ldots, s-1), \quad x_{ss} = 1,\\
&x_{sp} = \frac{1}{D_s}\left(k_{ps} - \sum_{q=1}^{s-1} D_q x_{qp} \overline{x}_{qs}\right) \quad (p = s+1, \ldots, n), \quad (76)\\
&f_{sp} = -x_{ps} - \sum_{q=p+1}^{s-1} x_{qs} f_{qp} \qquad (p = 1, \ldots, s-2),\\
&f_{s,\,s-1} = -x_{s-1,\,s},\quad f_{ss} = 1,\quad f_{sp} = 0\\
&\qquad (p = s+1, \ldots, n;\ s = 2, \ldots, n).
\end{aligned}$$

Если $D_s = 0$, то в соответствии с общим методом полагаем $x_{s1} = \ldots = x_{sn} = 0$.

Объем вычислений, необходимых для нахождения канонического разложения случайного вектора по формулам (76), в десятки, а в некоторых случаях в сотни раз меньше объема вычислений, необходимых для нахождения собственных значений и собственных векторов случайного вектора. При этом далеко не всегда приходится вычислять векторы f_p.

Предоставляем читателю самостоятельно показать, что формулы (76) можно получить, приняв за случайную величину V_1 первую координату вектора X^0, $V_1 = X_1^0$, а затем положив последовательно

$$X_s^0 = x_{s1} V_1 + \ldots + x_{s,\,s-1} V_{s-1} + V_s$$

и определив $x_{s1}, \ldots, x_{s,\,s-1}$ из условия некоррелированности величины V_s с $V_1, \ldots, V_{s-1}$.

Пример 15. В условиях примера 14 вычисление по формулам (76) с точностью до четырех знаков после запятой дает каноническое разложение случайного вектора X:

$$\begin{aligned}
X = {} & V_1\,[1\ \ 0{,}9278\ \ 0{,}8144\ \ 0{,}5722\ \ 0{,}3144]^T +\\
& + V_2\,[0\ \ 1\ \ 0{,}7135\ \ -0{,}6955\ \ -0{,}3740]^T +\\
& + V_3\,[0\ \ 0\ \ 1\ \ 0{,}4464\ \ 0{,}4597]^T +\\
& + V_4\,[0\ \ 0\ \ 0\ \ 1\ \ -0{,}2420]^T + V_5\,[0\ 0\ 0\ 0\ 1]^T,
\end{aligned}$$

дисперсии случайных величин $V_1, \ldots, V_5$:

$D_1 = 1{,}9400, \quad D_2 = 0{,}2299, \quad D_3 = 0{,}3362, \quad D_4 = 0{,}4867, \quad D_5 = 0{,}7765,$

и векторы $f_1, \ldots, f_5$:

$$f_1 = [1 \ 0 \ 0 \ 0 \ 0]^T,$$
$$f_2 = [-0{,}9278 \ 1 \ 0 \ 0 \ 0]^T,$$
$$f_3 = [-0{,}1525 \ -0{,}7135 \ 1 \ 0 \ 0]^T,$$
$$f_4 = [0{,}8991 \ 1{,}0140 \ -0{,}4464 \ 1 \ 0]^T,$$
$$f_5 = [-0{,}3737 \ 0{,}9474 \ -0{,}5677 \ 0{,}2420 \ 1]^T.$$

Проверка путем вычисления матрицы K_x по формуле (75) дает матрицу, полностью совпадающую с исходной матрицей K_x примера 14 с точностью до четырех знаков после запятой.

4.6. Геометрический смысл канонического разложения. Заметим теперь, что так как уравнение

$$u^T K_x u = \text{const}.$$

в случае матрицы K_x с действительными элементами представляет собой уравнение эллипсоида (эллипса при $n=2$) в силу неотрицательной определенности K_x, то задача нахождения канонического разложения действительного случайного вектора равносильна геометрической задаче нахождения сопряженных относительно эллипсоида направлений и приведения его уравнения к канонической форме (в общем случае в косоугольных координатах). В частном случае, приняв за оси координат оси эллипсоида, получаем прямоугольную систему координат, в которой уравнение эллипсоида имеет каноническую форму. Этой системе координат соответствует разложение случайного вектора по собственным векторам. Из этой геометрической интерпретации ясно, почему разложение случайного вектора на некоррелированные слагаемые называется каноническим.

Эллипсоид с уравнением $u^T K_x^{-1} u = 1$ (в случае матрицы K_x с действительными элементами) будем называть *эллипсоидом рассеивания.*

4.7. Построение случайного вектора с данной ковариационной матрицей. Докажем теперь утверждение § 3 о том, что любая неотрицательно определенная эрмитова матрица K может быть ковариационной матрицей (а следовательно, и моментом второго порядка) случайного вектора.

► Пусть $\lambda_1, \ldots, \lambda_n$ — собственные значения, $\varphi_1, \ldots, \varphi_n$ — соответствующие собственные векторы матрицы K, $U_1, \ldots, U_n$ — произвольные некоррелированные случай-

ные величины с нулевыми математическими ожиданиями и дисперсиями, равными соответственно $\lambda_1, \ldots, \lambda_n$ *). Рассмотрим случайный вектор

$$X=\sum_{k=1}^{n} U_p\varphi_p.$$

Его ковариационная матрица определяется формулой (67). А так как $\lambda_p\varphi_p=K\varphi_p$ и, следовательно, $\lambda_p\varphi_p^*=\varphi_p^*K$, то

$$K_x=\sum_{p=1}^{n}\lambda_p\varphi_p\varphi_p^*=\left(\sum_{p=1}^{n}\varphi_p\varphi_p^*\right)K.$$

Но

$$\sum_{p=1}^{n}\varphi_p\varphi_p^*=[\varphi_1\ \ldots\ \varphi_n]\begin{bmatrix}\varphi_1^*\\ \cdot\\ \cdot\\ \cdot\\ \varphi_n^*\end{bmatrix}=\Phi\Phi^*=I,$$

так как в силу (62)

$$\Phi^*\Phi=\begin{bmatrix}\varphi_1^*\\ \cdot\\ \cdot\\ \cdot\\ \varphi_n^*\end{bmatrix}[\varphi_1\ \ldots\ \varphi_n]=\begin{bmatrix}\varphi_1^*\varphi_1 & \ldots & \varphi_1^*\varphi_n\\ \cdot & \cdot\ \cdot\ \cdot & \cdot\\ \varphi_n^*\varphi_1 & \ldots & \varphi_n^*\varphi_n\end{bmatrix}=I,$$

и, следовательно, $\Phi^*=\Phi^{-1}$. Таким образом, $K_x=K$, что и доказывает утверждение § 3. ◀

§ 5. Другие числовые характеристики случайных величин

5.1. Моменты высших порядков. Моменты первого и второго порядка во многих случаях оказываются недостаточными. Поэтому для более полной характеристики случайных величин вводят моменты высших порядков. При

---

*) Например, можно взять случайные величины $U_p=U_p'+iU_p''$ $(p=1, \ldots, n)$ и определить совместную плотность величин U_1', U_1'', $\ldots$, U_n', U_n'' формулой

$$f(u_1', u_1'', \ldots, u_n', u_n'')=f_1(u_1', u_1'')\ldots f_n(u_n', u_n''),$$

где $f_1, \ldots, f_n$ — произвольные двумерные плотности, удовлетворяющие условиям $MU_p=0$, $DU_p=\lambda_p$ $(p=1, \ldots, n)$. В случае, когда матрица K имеет ранг $r<n$, $\lambda_{r+1}=\ldots=\lambda_n=0$ и случайные величины $U_{r+1}, \ldots, U_n$ с вероятностью 1 равны 0. В этом случае следует положить

$$f_p(u', u'')=\delta(u')\,\delta(u'')\qquad(p=r+1, \ldots, n).$$

этом обычно ограничиваются действительными величинами, что не приводит к потере общности (см. замечание в § 2.1).

Моментом порядка r (*r-м моментом*) скалярной случайной величины X называется математическое ожидание ее r-й степени:

$$\alpha_r = MX^r. \tag{77}$$

Центральным моментом порядка r случайной величины X называется r-й момент соответствующей центрированной случайной величины $X^0 = X - m_x$:

$$\mu_r = M(X^0)^r = M(X - m_x)^r. \tag{78}$$

Моментами порядка r случайного вектора X называются математические ожидания произведений $X_1^{k_1}, \ldots, X_n^{k_n}$, соответствующих всем целым неотрицательным $k_1, \ldots, k_n$, дающим в сумме r:

$$\alpha_{k_1, \ldots, k_n} = MX_1^{k_1} \ldots X_n^{k_n}, \quad k_1 + \ldots + k_n = r. \tag{79}$$

Центральными моментами порядка r случайного вектора X называются моменты порядка r соответствующего центрированного случайного вектора $X^0 = X - m_x$:

$$\mu_{k_1, \ldots, k_n} = M(X_1^0)^{k_1} \ldots (X_n^0)^{k_n}, \quad k_1 + \ldots + k_n = r. \tag{80}$$

Аналогично определяются моменты относительно любой данной точки. Моменты α относительно начала координат обычно называют *начальными*.

Пользуясь формулой бинома Ньютона и свойствами математических ожиданий, можно выразить начальные моменты через центральные, и наоборот. В частности, для моментов скалярной случайной величины получаем соотношения

$$\alpha_r = \sum_{p=0}^{r} C_r^p \mu_p m_x^{r-p}, \tag{81}$$

$$\mu_r = \sum_{p=0}^{r} (-1)^{r-p} C_r^p \alpha_p m_x^{r-p}. \tag{82}$$

Очевидно, что все моменты нулевого порядка равны 1. Момент первого порядка скалярной случайной величины X по определению представляет собой ее математическое ожидание, $\alpha_1 = m_x$, а моменты первого порядка случайного вектора X — математические ожидания его коорди-

нат. Все центральные моменты первого порядка равны нулю. Центральный момент второго порядка скалярной случайной величины X представляет собой ее дисперсию, а центральные моменты второго порядка случайного вектора X — элементы его ковариационной матрицы.

Кроме моментов первого и второго порядков, на практике иногда пользуются моментами третьего и четвертого порядков. Моментами высших порядков пользуются редко.

Центральный момент третьего порядка характеризует несимметричность распределения относительно математического ожидания. Поэтому за характеристику несимметричности распределения принимают безразмерную величину — отношение третьего центрального момента к кубу среднего квадратического отклонения:

$$\gamma_1 = \frac{\mu_3}{\sigma_x^3} = \frac{\mu_3}{D_x^{3/2}} = \frac{\mu_3}{\mu_2^{3/2}}. \tag{83}$$

Эта величина называется *асимметрией* или *скошенностью* распределения случайной величины.

Центральный момент четвертого порядка при данной дисперсии может служить характеристикой удельного веса больших отклонений от математического ожидания, а это в свою очередь определяет характер максимума в точке m_x симметричного распределения — «островершинность» или «плосковершинность» кривой распределения. Поэтому за характеристику поведения кривой распределения вблизи точки m_x принимают безразмерную величину

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3, \tag{84}$$

называемую *эксцессом* распределения случайной величины X. Смысл вычитания 3 из μ_4/μ_2^2 прояснится в § 6.

Пример 16. Для γ-распределения (пример 2.8) в силу (2.16)

$$\alpha_r = \frac{k^{\mu+1}}{\Gamma(\mu+1)} \int_0^\infty x^{r+\mu} e^{-kx}\, dx = \frac{\Gamma(r+\mu+1)}{k^r \Gamma(\mu+1)}.$$

В частном случае для показательного распределения ($\mu=0$) $\alpha_r = r!/k^r$.

Пример 17. Для β-распределения (пример 2.9)

$$\alpha_r = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^1 x^{r+p-1}(1-x)^{q-1}\, dx = \frac{\Gamma(p+q)\Gamma(p+r)}{\Gamma(p)\Gamma(p+q+r)}.$$

Пример 18. В условиях примера 2.14 моменты времени S работы системы после первого ремонта определяются формулой

$$\alpha_r = \int_0^{t_0} s^r \left[\lambda e^{-\lambda(t_0-s)} + e^{-\lambda t_0}\delta(s)\right] ds = \int_0^{t_0} s^r \lambda e^{-\lambda(t_0-s)}\, ds =$$

$$= t_0^r - \frac{r}{\lambda} t_0^{r-1} + \ldots + (-1)^{r-1} \frac{r!}{\lambda^{r-1}} t_0 + (-1)^r \frac{r!}{\lambda^r}\left(1 - e^{-\lambda t_0}\right).$$

Пример 19. В условиях примера 2.16

$$\alpha_r = \int_0^{2l} u^r \left[f(u+a-l) + f(b+l-u)\right] du + (2l)^r p_2.$$

5.2. Квантили. Моменты случайных величин удобны, когда все необходимые моменты (практически не выше четвертого порядка) существуют. Однако, как показывает пример 6, случайная величина может не иметь моментов. Поэтому для скалярных случайных величин иногда вводят другие числовые характеристики, связанные со значениями функции распределения.

За характеристику положения распределения на числовой оси (среднее значение) принимают точку $\zeta_{1/2}$, в которой функция распределения переходит от значений, меньших 1/2, к значениям, большим 1/2: $F(\zeta_{1/2}) \leqslant 1/2$, $F(\zeta_{1/2}+0) \geqslant 1/2$. Такая точка $\zeta_{1/2}$ называется *медианой* случайной величины. Медиана всегда существует и может даже быть неоднозначно определенной. Действительно, функция распределения изменяется монотонно от 0 до 1. Поэтому всегда найдется такая точка $\zeta_{1/2}$, в которой она скачком или непрерывно переходит через значение 1/2. Если же существует интервал (α, β), на котором $F(x) = 1/2$, то любая точка этого интервала может служить значением медианы.

Для характеристики разброса значений случайной величины пользуются точками, в которых функция распределения переходит через другие значения. Точка ζ_p, в которой функция распределения переходит от значений, меньших p, к значениям, большим p, $p \in (0, 1)$, называется *квантилью порядка* p: $F(\zeta_p) \leqslant p$, $F(\zeta_p+0) \geqslant p$. Все квантили существуют у любой действительной скалярной случайной величины, и некоторые из них могут быть определены неоднозначно. Это доказывается совершенно так же, как существование медианы.

Чаще всего пользуются, кроме медианы $\zeta_{1/2}$, квантилями $\zeta_{1/4}$ и $\zeta_{3/4}$, которые называются *квартилями*. Величина $E=(\zeta_{3/4}-\zeta_{1/4})/2$ при этом принимается за характеристику разброса значений случайной величины и называется *срединным* или *вероятным отклонением* или *семиинтерквартильной широтой* распределения случайной величины. Иногда пользуются квантилями $\zeta_{0,1}, \zeta_{0,2}, \ldots$ $\ldots, \zeta_{0,9}$, которые называются *децилями*.

Легко видеть, что с помощью квантилей можно характеризовать распределение с любой степенью точности, если взять их достаточно много.

Недостаток квантилей состоит в том, что для них не существует простых формул типа (52) для перехода от случайных величин к линейным функциям этих случайных величин, в частности к их суммам.

§ 6. Одномерное нормальное распределение

6.1. Моменты. Изучим теперь нормальное распределение на числовой оси (пример 2.6). Это распределение определяется плотностью

$$f(x)=\sqrt{\frac{c}{2\pi}}e^{-\frac{1}{2}c(x-a)^2}, \qquad c>0. \tag{85}$$

На рис. 15 показана нормальная кривая распределения, изображающая эту плотность. Эта кривая, как видно из (85), симметрична относительно точки a, в которой плотность достигает максимума. Коэффициент при показательной функции определен из условия равенства единице интеграла от плотности, распространенного на всю числовую ось.

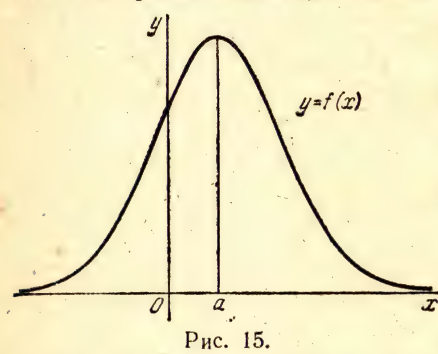


Рис. 15.

Чтобы вывести выражение для этого коэффициента, обозначим его A. Тогда получим

$$A\int_{-\infty}^{\infty} e^{-\frac{1}{2}c(x-a)^2}\,dx=1.$$

Заменой переменных $x=a+t\sqrt{2/c}$ интеграл в левой части приводится к известному интегралу Пуассона:

$$A\sqrt{\frac{2}{c}}\int_{-\infty}^{\infty} e^{-t^2}\,dt=1.$$

Отсюда, принимая во внимание, что (приложение 2)

$$\int_{-\infty}^{\infty} e^{-t^2}\,dt=\sqrt{\pi}, \qquad (86)$$

находим $A=\sqrt{c/2\pi}$.

Из симметрии нормальной кривой распределения относительно точки $x=a$ непосредственно следует, что математическое ожидание случайной величины X, распределенной по нормальному закону, равно параметру a в выражении нормальной плотности, $m_x=a$. К этому же результату приводит и формальное вычисление m_x по формуле (3):

$$m_x=\sqrt{\frac{c}{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{1}{2}c(x-a)^2}\,dx=$$

$$=\frac{a}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-t^2}dt+\sqrt{\frac{2}{c\pi}}\int_{-\infty}^{\infty} te^{-t^2}\,dt=a$$

в силу формулы (86) и равенства нулю последнего интеграла как интеграла от нечетной функции в симметричных относительно начала координат пределах.

Таким образом, *параметр a в выражении* (85) *одномерной нормальной плотности представляет собой математическое ожидание случайной величины.*

Центральные моменты нормально распределенной случайной величины X определяются согласно (78) и (6) формулой

$$\mu_r=\sqrt{\frac{c}{2\pi}}\int_{-\infty}^{\infty} (x-a)^r e^{-\frac{1}{2}c(x-a)^2}\,dx=$$

$$=\frac{1}{\sqrt{\pi}}\left(\frac{2}{c}\right)^{r/2}\int_{-\infty}^{\infty} t^r e^{-t^2}\,dt.$$

Отсюда видно, что все нечетные центральные моменты нормально распределенной случайной величины равны

нулю. Для центральных моментов четных порядков получаем

$$\mu_{2p} = \frac{1}{\sqrt{\pi}} \left(\frac{2}{c}\right)^p \int_{-\infty}^{\infty} t^{2p} e^{-t^2}\, dt = \frac{2}{\sqrt{\pi}} \left(\frac{2}{c}\right)^p \int_{0}^{\infty} t^{2p} e^{-t^2}\, dt. \quad (87)$$

Интегрируя по частям, имеем

$$\mu_{2p} = \frac{1}{\sqrt{\pi}} \left(\frac{2}{c}\right)^p \int_{0}^{\infty} t^{2p-1} e^{-t^2} dt^2 = \frac{1}{\sqrt{\pi}} \left(\frac{2}{c}\right)^p \left[- t^{2p-1} e^{-t^2}\right]\Big|_0^{\infty} +$$

$$+ \frac{2p-1}{\sqrt{\pi}} \left(\frac{2}{c}\right)^p \int_{0}^{\infty} t^{2p-2} e^{-t^2}\, dt = \frac{2p-1}{c} \mu_{2p-2}.$$

Таким образом, мы получили рекуррентную формулу

$$\mu_{2p} = \frac{2p-1}{c} \mu_{2p-2}. \quad (88)$$

Положив здесь $p=1$ и имея в виду, что все моменты нулевого порядка равны 1, находим дисперсию нормально распределенной случайной величины X:

$$D_x = \mu_2 = 1/c.$$

Таким образом, *параметр c в выражении* (85) *нормальной плотности представляет собой величину, обратную дисперсии случайной величины*, $c = D_x^{-1} = \mu_2^{-1}$.

Для получения общей формулы для четных центральных моментов перепишем (88) в виде

$$\mu_{2p} = (2p-1)\, \mu_2 \mu_{2p-2}.$$

Положив здесь последовательно $p = 2, 3, \ldots, k$ и перемножив полученные равенства почленно, получим после сокращений

$$\mu_{2k} = 1 \cdot 3 \cdot \ldots \cdot (2k-1)\, \mu_2^k = (2k-1)!!\, \mu_2^k. \quad (89)$$

Эта формула выражает все четные центральные моменты нормально распределенной случайной величины X через ее дисперсию $D_x = \mu_2$. В частности, при $k=2$ формула (89) определяет центральный момент четвертого порядка $\mu_4 = 3\mu_2^2$. Подставив это выражение в (84), убеждаемся в том, что эксцесс нормального распределения равен нулю.

Нормальное распределение широко распространено в природе. В большинстве задач практики распределение

случайной величины можно считать нормальным. Поэтому нормальное распределение обычно принимается за эталон для сравнения распределений. Асимметрия и эксцесс вводятся для того, чтобы характеризовать отклонение распределения от нормального. Поэтому их определяют так, чтобы для нормального распределения они были равны нулю. Этим и объясняется ввод слагаемого —3 в определение (84) эксцесса.

Учитывая, что $a = m_x$, $c = D_x^{-1} = \sigma_x^{-2}$, выражение (85) для одномерной нормальной плотности часто записывают в виде

$$f(x) = \frac{1}{\sqrt{2\pi D}} e^{-\frac{(x-m)^2}{2D}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \qquad (90)$$

опуская для краткости индексы у математического ожидания $m = m_x$, дисперсии $D = D_x$ и среднего квадратического отклонения $\sigma = \sigma_x$.

Формула (90) показывает, что *нормальное распределение полностью определяется моментами первого и второго порядков*. Зная математическое ожидание и дисперсию нормально распределенной случайной величины, можно найти ее плотность.

6.2. Вероятность попадания в интервал. Выведем теперь формулу для вероятности попадания нормально распределенной случайной величины X в данный интервал (α, β).

Подставив выражение (90) плотности в общую формулу (2.10) для вероятности попадания в интервал, получим

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx.$$

После замены переменных $x = m + \sigma z$, $z = (x - m)/\sigma$ будем иметь

$$P(\alpha < X < \beta) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-m}{\sigma}}^{\frac{\beta-m}{\sigma}} e^{-z^2/2} dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\beta-m}{\sigma}} e^{-z^2/2} dz - \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\alpha-m}{\sigma}} e^{-z^2/2} dz.$$

Неопределенный интеграл от функции $e^{-z^2/2}$ не выражается в элементарных функциях. Поэтому для вычисления интегралов в предыдущей формуле вводят новую функцию

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-z^2/2} dz. \tag{91}$$

Для этой функции, называемой *функцией Лапласа*, составлены таблицы (приложение 3, табл. 1) *). При помощи функции Лапласа предыдущая формула для вероятности попадания нормально распределенной случайной величины в интервал (α, β) может быть представлена в виде

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m_x}{\sigma_x}\right) - \Phi\left(\frac{\alpha - m_x}{\sigma_x}\right). \tag{92}$$

В этой формуле явно показано, что параметры m и σ представляют собой соответственно математическое ожидание m_x и среднее квадратическое отклонение σ_x случайной величины X **).

Докажем теперь, что функция Лапласа нечетная. Заменив в (91) u на $-u$, будем иметь

$$\Phi(-u) = \frac{1}{\sqrt{2\pi}} \int_0^{-u} e^{-z^2/2} dz.$$

Сделав замену переменных $z = -s$, получим

$$\Phi(-u) = -\frac{1}{\sqrt{2\pi}} \int_0^u e^{-s^2/2} ds = -\Phi(u),$$

что и доказывает нечетность функции $\Phi(u)$. Вследствие этого в таблицах даются ее значения только для положительных значений u.

6.3. Случай симметричного интервала. В частном случае, когда интервал (α, β) симметричен относительно математического ожидания, $\alpha = m_x - \varepsilon$, $\beta = m_x + \varepsilon$,

*) В дальнейшем при ссылках на таблицы всегда имеются в виду таблицы, приведенные в приложении 3.

**) В разных руководствах функцию Лапласа определяют поразному. Различные определения отличаются одно от другого масштабом аргумента, масштабом самой функции и пределами интеграла. Соответственно изменяются и формулы для вероятности попадания.

формула (92) дает

$$P(m_x - \varepsilon < X < m_x + \varepsilon) = P(|X - m_x| < \varepsilon) = \\ = \Phi\left(\frac{\varepsilon}{\sigma_x}\right) - \Phi\left(-\frac{\varepsilon}{\sigma_x}\right).$$

Отсюда, принимая во внимание нечетность функции $\Phi(u)$, получаем

$$P(|X - m_x| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma_x}\right). \quad (93)$$

Пример 20. Полагая в формуле (93) последовательно $\varepsilon = \sigma_x$, $2\sigma_x$, $3\sigma_x$, $4\sigma_x$ и пользуясь таблицей функции Лапласа (приложения 3, табл. 1), находим

$$P(|X - m_x| < \sigma_x) = 2\Phi(1) \approx 0{,}683 \approx 0{,}68,$$
$$P(|X - m_x| < 2\sigma_x) = 2\Phi(2) \approx 0{,}954 \approx 0{,}95,$$
$$P(|X - m_x| < 3\sigma_x) = 2\Phi(3) \approx 0{,}997,$$
$$P(|X - m_x| < 4\sigma_x) = 2\Phi(4) \approx 0{,}999994.$$

Таким образом, с вероятностью около 68% значения нормально распределенной случайной величины отклоняются от ее математического ожидания не больше, чем на одно среднее квадратическое отклонение, с вероятностью 95% — не больше чем на два средних квадратических отклонения и с вероятностью 99,7% — не больше чем на три средних квадратических отклонения.

Мы видим, что нормально распределенная случайная величина лишь с вероятностью 0,3% может отклониться от своего математического ожидания больше чем на три средних квадратических отклонения. Иными словами, в среднем лишь в трех опытах из тысячи, ее отклонение от математического ожидания будет больше $3\sigma_x$. Поэтому обычно считают областью практически возможных значений нормально распределенной случайной величины интервал $(m_x - 3\sigma_x, m_x + 3\sigma_x)$.

Положив в (92) $\alpha = -\infty$, $\beta = x$ и приняв во внимание, что $\Phi(\infty) = 1/2$, находим функцию распределения нормально распределенной случайной величины:

$$F(x) = \Phi\left(\frac{x - m_x}{\sigma_x}\right) + \frac{1}{2}\ ^{*)}. \quad (94)$$

6.4. Квантили. Из симметрии нормального распределения непосредственно следует, что медиана совпадает

---

*) Для вычисления нормальной функции распределения и плотности в [94] служит программа NDTR. Пользуясь этой программой, можно вычислять на ЭВМ вероятности попадания нормально распределенных случайных величин в любые интервалы.

с математическим ожиданием, $\zeta_{1/2}=m_x=a$, и что $\zeta_{1/2}-\zeta_{1/4}=\zeta_{3/4}-\zeta_{1/2}=E_x$, где E_x — срединное отклонение величины X. Для определения срединного отклонения E_x достаточно положить в формуле (94) $x=\zeta_{3/4}=m_x+E_x$ и приравнять полученное выражение 3/4. Тогда получим уравнение $\Phi\left(\frac{E_x}{\sigma_x}\right)=\frac{1}{4}$. Корень этого уравнения обычно обозначается $\rho\sqrt{2}$ $(\rho=0{,}4769\ldots)$. Таким образом, $\frac{E_x}{\sigma_x}=\rho\sqrt{2}$, откуда находим срединное отклонение

$$E_x=\rho\sqrt{2}\,\sigma_x\approx 0{,}674\sigma_x.$$

Глава 4

ПРОЕКЦИИ СЛУЧАЙНЫХ ВЕКТОРОВ И ИХ РАСПРЕДЕЛЕНИЯ

§ 1. Распределения проекций случайного вектора

1.1. Проекции вектора. Любой вектор $x=\{x_1, \ldots, x_N\}$ в N-мерном пространстве можно однозначно представить в виде

$$x=\sum_{k=1}^{N} x_k e_k, \tag{1}$$

где $e_1, \ldots, e_N$ — единичные векторы осей координат,

$$e_1=\{1, 0, \ldots, 0\},\ e_2=\{0, 1, 0, \ldots, 0\},\ \ldots,\ e_N=\{0, \ldots, 0, 1\}.$$

Выберем какие-нибудь n из N векторов $e_1, \ldots, e_N$, скажем $e_{r_1}, \ldots, e_{r_n}$, $r_1 < r_2 < \ldots < r_n$, $n < N$. *Подпространством, образованным векторами* $e_{r_1}, \ldots, e_{r_n}$, называется множество всех векторов x N-мерного пространства, у которых отличны от нуля только координаты $x_{r_1}, \ldots, x_{r_n}$, а остальные равны нулю. Два подпространства, одно из которых образовано каким-нибудь n из векторов $e_1, \ldots, e_N$, а другое — всеми оставшимися, называются *взаимно дополнительными*. Например, подпространство, образованное векторами $e_1, \ldots, e_n$, и подпространство, образованное векторами $e_{n+1}, \ldots, e_N$, взаимно дополнительны.

Проекцией вектора x на подпространство, образованное векторами $e_{r_1}, \ldots, e_{r_n}$, называется вектор x', полученный из x заменой всех его координат, кроме $x_{r_1}, \ldots, x_{r_n}$, нулями:

$$x'=\sum_{k=1}^{n} x_{r_k} e_{r_k}. \tag{2}$$

В частности, проекцией вектора x на k-ю координатную ось — одномерное подпространство, образованное вектором e_k, является вектор $x_k e_k$. Ввиду того, что проекция вектора x на k-ю координатную ось полностью определяется одним числом x_k, мы в дальнейшем не будем

делать различия между проекцией вектора на ось и соответствующей его координатой, разумеется, если это не может вызвать путаницу.

Пусть $s_1, \ldots, s_{N-n}$ — оставшиеся числа из множества $\{1, \ldots, N\}$ после выбора из него чисел $r_1, \ldots, r_n$, расположенные в порядке возрастания, $s_1 < \ldots < s_{N-n}$. Вектор x', определяемый формулой (2), и вектор

$$x'' = \sum_{h=1}^{N-n} x_{s_h} e_{s_h} \tag{3}$$

представляют собой проекции вектора x на два взаимно дополнительных подпространства. Сумма этих двух проекций равна вектору x, $x' + x'' = x$, так как сумма (3) содержит все слагаемые суммы (1), которые не входят в сумму (2), и не содержит ни одного слагаемого суммы (2).

Если X — N-мерный случайный вектор, то его проекция на подпространство, образованное единичными векторами $e_{r_1}, \ldots, e_{r_n}$, представляет собой n-мерный случайный вектор, а его проекция на дополнительное подпространство — $(N-n)$-мерный случайный вектор.

Так как изменением нумерации векторов $e_1, \ldots, e_N$ (т. е. их порядка) всегда можно любые n из них сделать n первыми, то в дальнейшем, рассматривая проекцию N-мерного вектора на n-мерное подпространство, мы всегда будем считать, что это подпространство образовано единичными векторами $e_1, \ldots, e_n$. Тогда дополнительное подпространство будет образовано единичными векторами $e_{n+1}, \ldots, e_N$. При этом для простоты будем полагать $N = n + m$ и обозначать $(n+m)$-мерный вектор буквой z, а его проекции на n-мерное и дополнительное m-мерное подпространства — соответственно буквами x и y. Иными словами, будем полагать $z_1 = x_1, \ldots, z_n = x_n$, $z_{n+1} = y_1, \ldots, z_{n+m} = y_m$ и $x = \{x_1, \ldots, x_n\}$, $y = \{y_1, \ldots, y_m\}$.

1.2. Функция распределения проекции случайного вектора. Предположим, что известна функция распределения $F(x, y)$ двумерного случайного вектора Z с координатами (проекциями на оси координат) X, Y, и поставим задачу определить распределения случайных величин X и Y, рассматриваемых по отдельности.

▶ В силу определения (2.29) и соглашения § 2.2 о записи неравенств между векторами

$$F(x, y) = P(\{X < x\}\{Y < y\}). \tag{4}$$

Для нахождения функции распределения $F_1(x)$ случайной величины X достаточно заметить, что выполнение неравенства $X < x$ равноценно совместному выполнению этого неравенства и достоверного неравенства $Y < \infty$. Поэтому

$$F_1(x) = P(X < x) = P(\{X < x\}\{Y < \infty\}). \qquad (5)$$

Из сравнения этой формулы с (4) видно, что

$$F_1(x) = F(x, \infty). \blacktriangleleft \qquad (6)$$

Совершенно так же определяется функция распределения $F_2(y)$ случайной величины Y:

$$F_2(y) = F(\infty, y). \qquad (7)$$

Таким образом, *чтобы найти функцию распределения одной координаты двумерного случайного вектора, следует положить переменную, соответствующую другой координате, равной* ∞.

1.3. Плотность проекции случайного вектора. Найдем теперь плотность случайной величины X.

► Применив для вычисления вероятности в (5) формулу (2.9) и изменив обозначение переменной интегрирования x на u, получим

$$F_1(x) = \int_{-\infty}^{x} du \int_{-\infty}^{\infty} f(u, y)\, dy = \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} f(u, y)\, dy \right\} du.$$

Сравнив эту формулу с (2.30), получаем следующую формулу для плотности $f_1(x)$ случайной величины X:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy. \blacktriangleleft \qquad (8)$$

Совершенно так же получаем формулу для плотности $f_2(y)$ случайной величины Y:

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx. \qquad (9)$$

Таким образом, *чтобы найти плотность одной координаты двумерного случайного вектора, следует проинтегрировать его плотность по переменной, соответствующей другой координате.*

Формулы (6) — (9) справедливы и в том случае, когда X и Y представляют собой любые случайные векторы. Если X представляет собой n-мерный случайный вектор с координатами $X_1, \ldots, X_n$, а Y — m-мерный случайный вектор с координатами $Y_1, \ldots, Y_m$, то $f(x, y)$ является плотностью составного $(n+m)$-мерного случайного вектора с координатами $X_1, \ldots, X_n, Y_1, \ldots, Y_m$, т. е. совместной плотностью $n+m$ случайных величин $X_1, \ldots, X_n, Y_1, \ldots, Y_m$; интеграл по переменной y в (8) следует понимать как m-кратный интеграл по координатам вектора y, а интеграл по переменной x в (9) — как n-кратный интеграл по координатам вектора x.

Таким образом, формулы (6) — (9) определяют распределения проекций любого случайного вектора на взаимно дополнительные подпространства.

► Формула (8) дает возможность доказать сделанное в § 3.1 утверждение, что математическое ожидание проекции (в частности, одной координаты) случайного вектора можно вычислить по формуле (3.3), рассматривая эту проекцию как функцию случайного вектора. Действительно, рассматривая величину X как функцию случайного вектора $\{X, Y\}$, по формуле (3.3) получаем

$$m_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y)\, dx\, dy = \int_{-\infty}^{\infty} x\, dx \int_{-\infty}^{\infty} f(x, y)\, dy,$$

или в силу (8)

$$m_x = \int_{-\infty}^{\infty} x f_1(x)\, dx.$$

То же получается по формуле (3.6). ◄

Пример 1. Плотность двумерного нормально распределенного случайного вектора определяется формулой (2.22) примера 2.11 при $n=2$:

$$f(x, y) = \frac{\sqrt{c_{11}c_{22} - c_{12}^2}}{2\pi} \exp\left\{-\frac{1}{2}\left[c_{11}(x-a)^2 + 2c_{12}(x-a)(y-b) + c_{22}(y-b)^2\right]\right\}. \quad (10)$$

Для определения плотности проекции этого случайного вектора на ось x подставим выражение (10) в (8):

$$f_1(x) = \frac{\sqrt{c_{11}c_{12} - c_{12}^2}}{2\pi} e^{-\frac{1}{2} c_{11}(x-a)^2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} c_{22}(y-b)^2 - \right.$$

$$\left. - c_{12}(x-a)(y-b)\right\} dy =$$

$$= \frac{\sqrt{c_{11}c_{12} - c_{12}^2}}{2\pi \sqrt{c_{22}}} e^{-\frac{1}{2} c_{11}(x-a)^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2 - \frac{c_{12}}{\sqrt{c_{22}}}(x-a)t} dt.$$

Воспользовавшись для вычисления интеграла формулой (приложение 2)

$$\int_{-\infty}^{\infty} e^{\eta t - \frac{1}{2}t^2} dt = \sqrt{2\pi}\, e^{\frac{1}{2}\eta^2}, \tag{11}$$

получаем

$$f_1(x) = \sqrt{\frac{c_{11}c_{22} - c_{12}^2}{2\pi c_{22}}}\, e^{-\frac{1}{2}\frac{c_{11}c_{22} - c_{12}^2}{c_{22}}(x-a)^2}. \tag{12}$$

Сравнив эту формулу с (3.90), видим, что проекция X двумерного нормально распределенного случайного вектора $\{X, Y\}$ на ось x распределена нормально, причем ее математическое ожидание равно a, $m_x = a$, а дисперсия определяется формулой

$$D_x = \frac{c_{22}}{c_{11}c_{22} - c_{12}^2}. \tag{13}$$

По симметрии заключаем, что проекция Y вектора $\{X, Y\}$ на ось y распределена нормально, причем ее математическое ожидание равно b, $m_y = b$, а дисперсия определяется формулой

$$D_y = \frac{c_{11}}{c_{11}c_{22} - c_{12}^2}. \tag{14}$$

Вычислим еще ковариацию величин X и Y. Полагая для краткости $x - m_x = x - a = u$, $y - m_y = y - b = v$, по формуле (3.22) находим

$$k_{xy} = \frac{\sqrt{c_{11}c_{22} - c_{12}^2}}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv e^{\frac{1}{2}(c_{11}u^2 + 2c_{12}uv + c_{22}v^2)} du\, dv =$$

$$= \frac{\sqrt{c_{11}c_{22} - c_{12}^2}}{2\pi c_{22}} \int_{-\infty}^{\infty} u e^{-\frac{1}{2}c_{11}u^2} du \int_{-\infty}^{\infty} t e^{-\frac{c_{12}}{\sqrt{c_{22}}}ut - \frac{1}{2}t^2} dt.$$

Применив для вычисления интеграла по t формулу

$$\int_{-\infty}^{\infty} t e^{\eta t - \frac{1}{2}t^2} dt = \sqrt{2\pi}\, \eta e^{\frac{1}{2}\eta^2},$$

которая легко выводится дифференцированием формулы (11) по параметру η, получаем

$$k_{xy}=-\frac{c_{12}}{c_{22}}\sqrt{\frac{c_{11}c_{22}-c_{12}^2}{2\pi c_{22}}}\int_{-\infty}^{\infty}u^2e^{-\frac{1}{2}\frac{c_{11}c_{22}-c_{12}^2}{c_{22}}u^2}\,du.$$

Последний интеграл вместе с квадратным корнем перед ним согласно (3.16) и (12) представляет собой дисперсию D_x величины X, которая определяется формулой (13). Следовательно,

$$k_{xy}=-\frac{c_{12}}{c_{11}c_{22}-c_{12}^2}. \tag{15}$$

Формулы (13)—(15) показывают, что ковариационная матрица двумерного нормально распределенного случайного вектора и матрица C коэффициентов квадратичной формы в выражении (10) плотности являются взаимно обратными матрицами. В § 4 мы увидим, что это справедливо и для случайного вектора любой размерности.

Пример 2. Предлагаем читателю самостоятельно найти плотность проекции n-мерного нормально распределенного случайного вектора на $(n-1)$-мерное подпространство, образованное какими-нибудь $n-1$ осями координат.

Пример 3. Показать, что если двумерный вектор распределен равномерно на прямоугольнике (пример 2.10), то его проекция распределена равномерно на соответствующей проекции этого прямоугольника. Обобщить этот результат на равномерно распределенные векторы на параллелепипедах в любых конечномерных пространствах.

Пример 4. Для случайного вектора, равномерно распределенного внутри эллипса (пример 2.10), формула (8) дает

$$f_1(x)=\int\limits_{-b\sqrt{1-\frac{x^2}{a^2}}}^{b\sqrt{1-\frac{x^2}{a^2}}}\frac{dy}{\pi ab}=\frac{2}{\pi a}\sqrt{1-\frac{x^2}{a^2}},\quad f_1(x)=0\ \text{ при }\ |x|\geqslant a.$$

В данном случае, несмотря на равномерное распределение случайного вектора на плоскости, его координаты не распределены равномерно. Это и понятно: при каждом значении x вероятность попадания в бесконечно малый интервал длины Δx пропорциональна длине хорды эллипса, проходящей через точку x перпендикулярно оси x (рис. 16). Следовательно, и плотность величины X пропорциональна длине этой хорды. Отсюда ясно, что кривая распределения величины X на интервале $(-a, a)$ представляет собой половину эллипса (рис. 17).

Предлагаем читателю самостоятельно показать, что полученная формула определяет распределение проекции случайного вектора на ось x и в более сложном случае вектора, равномерно распределенного внутри эллипса, оси которого не совпадают с осями координат (пример 3.9). Найти для этого случая также распределение проекции вектора на ось y.

Пример 5. Рассмотрим случай, когда величина Y является определенной функцией случайной величины X, $Y=\varphi(X)$. В этом

случае, как было показано в п. 2.3.4, совместная плотность случайных величин X и Y определяется формулой

$$f(x, y) = g(x)\,\delta(y - \varphi(x)).$$

Подставив это выражение в (8), находим плотность величины X:

$$f_1(x) = \int_{-\infty}^{\infty} g(x)\,\delta(y - \varphi(x))\,dy = g(x).$$

Эта формула доказывает высказанное в п. 2.3.4 утверждение, что множитель $g(x)$ представляет собой плотность случайной величины X.

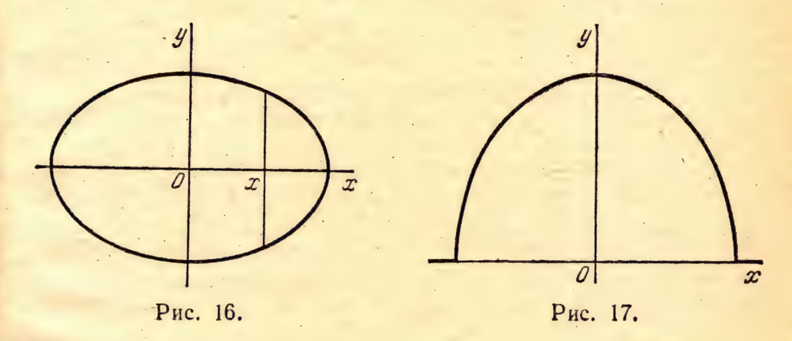


Рис. 16. Рис. 17.

И это верно как в случае скалярных, так и в случае векторных величин X и Y.

Точно так же по формуле (9) находим плотность величины Y:

$$f_2(y) = \int_{-\infty}^{\infty} g(x)\,\delta(y - \varphi(x))\,dx.$$

Для вычисления этого интеграла следует разбить область, где $g(x) \neq 0$, на интервалы монотонности функции $\varphi(x)$ и на каждом таком интервале сделать замену переменных $\eta = \varphi(x)$, после чего останется выполнить интегрирование по η, пользуясь определением интеграла от δ-функции (формула (5) приложения 1). Мы выполним здесь это вычисление для частного случая примера 3.10, когда $\varphi(x) = x^2$, $g(x) = (1/2a)1(a - |x|)$. В этом случае

$$f_2(y) = \frac{1}{2a}\int_{-a}^{a} \delta(y - x^2)\,dx = \frac{1}{2a}\left[\int_{-a}^{0} \delta(y - x^2)\,dx + \int_{0}^{a} \delta(y - x^2)\,dx\right].$$

Сделав замену переменных $\eta = x^2$, будем иметь в первом интеграле $x = -\sqrt{\eta}$, а во втором $x = +\sqrt{\eta}$. В результате получим

$$f_2(y) = \frac{1}{2a}\int_{0}^{a^2} \delta(y - \eta)\,\frac{d\eta}{\sqrt{\eta}} = \frac{1}{2a\sqrt{y}}1(a^2 - y).$$

Таким образом, обе координаты случайного вектора являются в данном случае непрерывными случайными величинами, несмотря на то что его плотность содержит множителем δ-функцию. Это иллюстрирует сказанное в п. 2.3.4.

Пример 6. В случае вырожденного распределения в n-мерном пространстве, подставив выражение (2.28) плотности в (8) с заменой x и y соответственно на x' и x'', найдем плотность случайного вектора X' (проекции вектора X на подпространство, образованное m первыми единичными векторами $e_1, \ldots, e_m$):

$$f_2(x') = \int_{-\infty}^{\infty} g(x')\, \delta(x'' - A(x' - c') - c'')\, dx'' = g(x').$$

§ 2. Условные распределения проекций случайного вектора

2.1. Условная плотность проекции случайного вектора. В задачах практики, имея дело со случайными векторами, часто приходится определять распределение одних координат случайного вектора при известных значениях других — так называемые условные распределения. С такой задачей мы встречаемся, например, когда интересующая нас случайная величина X непосредственно не наблюдаема, но может наблюдаться некоторая другая величина Y. Например, при любом измерении мы наблюдаем не саму измеряемую величину X, а лишь результат измерения Y, представляющий собой сумму измеряемой величины и ошибки измерения. В таких случаях, определив из опыта значение y величины Y, приходится искать соответствующее условное распределение величины X. Основная трудность при нахождении таких условных распределений состоит в том, что событие B вида $Y = y$, относительно которого приходится определять условные вероятности, имеет вероятность, равную нулю, в то время как определение (1.16) условной вероятности применимо только в случае, когда $P(B) \neq 0$.

Рассмотрим сначала непрерывный двумерный случайный вектор с плотностью $f(x, y)$. В этом случае $P(Y = y) = 0$. Поэтому формула (1.16) не определяет условные вероятности событий при условии $Y = y$. В то же время интуитивно ясно, что условная плотность величины X при данном y существует и пропорциональна плотности $f(x, y)$ при данном фиксированном значении y (условная кривая распределения величины X при данном y изображается в некотором масштабе соответствующим сечением

поверхности распределения $z=f(x, y)$). Из этих соображений, принимая во внимание свойство (2.11) плотности и формулу (9), получаем следующее выражение условной плотности величины X при данном значении y величины Y:

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)}. \tag{16}$$

Аналогично находим условную плотность величины Y при данном значении x величины X:

$$f_2(y \mid x) = \frac{f(x, y)}{f_1(x)}. \tag{17}$$

► Чтобы дать строгое определение условного распределения одной координаты случайного вектора при данном значении другой, найдем сначала условное распределение величины X относительно события $y_1 \leqslant Y < y_2$, имеющего отличную от нуля вероятность. По формуле (1.16) находим условную функцию распределения:

$$F_{y_1, y_2}(x) = P(X < x \mid y_1 \leqslant Y < y_2) = \frac{P(\{X < x\}\{y_1 \leqslant Y < y_2\})}{P(y_1 \leqslant Y < y_2)}.$$

Применив для вычисления вероятностей формулы (2.9) и (2.10), получим

$$F_{y_1, y_2}(x) = \frac{\int\limits_{-\infty}^{x} du \int\limits_{y_1}^{y_2} f(u, v)\, dv}{\int\limits_{y_1}^{y_2} f_2(v)\, dv}. \blacktriangleleft \tag{18}$$

Положим теперь $y_1 = y$, $y_2 = y + \Delta y$ и определим *условное распределение случайной величины X при данном значении y величины Y*, имеющем нулевую вероятность, как предел условного распределения X относительно события $y \leqslant Y < y + \Delta y$ при $\Delta y \to 0$.

► Чтобы осуществить предельный переход, перепишем формулу (18) в виде

$$F_{y_1, y_2}(x) = \frac{\int\limits_{y_1}^{y_2} dv \int\limits_{-\infty}^{x} f(u, v)\, du}{\int\limits_{y_1}^{y_2} f_2(v)\, dv}. \tag{19}$$

Если интеграл

$$\int_{-\infty}^{x} f(u, v)\, du$$

при любом x и функция $f_2(v)$ — непрерывные функции v в точке $v = y$, то они непрерывны также и в некоторой окрестности точки y. Поэтому, положив $y_1 = y$, $y_2 = y + \Delta y$, можно при достаточно малом Δy применить к интегралам в (19) теорему о среднем значении. В результате будем иметь

$$F_{y,\, y+\Delta y}(x) = \frac{\int_{-\infty}^{x} f(u, \eta)\, du}{f_2(\eta')}, \quad \eta,\ \eta' \in (y,\ y + \Delta y).$$

Так как $\eta, \eta' \to y$ при $\Delta y \to 0$, то правая часть полученного равенства стремится к определенному пределу при $\Delta y \to 0$, если $f_2(y) \neq 0$. Следовательно, в этом случае существует и предел левой части, который по определению представляет собой условную функцию распределения $F_1(x \mid y)$ случайной величины X при данном значении y величины Y. Таким образом, получаем

$$F_1(x \mid y) = \lim_{\Delta y \to 0} F_{y,\, y+\Delta y}(x) = \frac{\int_{-\infty}^{x} f(u, y)\, du}{f_2(y)}. \qquad (20)$$

Сравнив правую часть с выражением (2.30) функции распределения через плотность, получаем формулу (16) для условной плотности $f_1(x \mid y)$ величины X при $Y = y$. ◀

Таким образом, мы доказали, что *при всех значениях* y, *при которых* $f_2(y)$ *и интеграл в* (20) *при любом* x *являются непрерывными функциями* y *и* $f_2(y) \neq 0$, *условная плотность величины* X *при данном значении* y *величины* Y *определяется формулой* (16). При $f_2(y) = 0$ условное распределение величины X при $Y = y$ остается неопределенным (не существует).

Отметим, что условия, при которых была выведена формула (20), могут выполняться и в том случае, когда $f(x, y)$ содержит δ-функции (примеры 5 и 6).

▶ Для определения условного распределения величины X при значении y_s величины Y, имеющем отличную от нуля вероятность p_s, заметим, что событие $Y = y_s$ (попадание в прямую $y = y_s$) может иметь отличную от

нуля вероятность только в том случае, когда $f(x, y)$ содержит слагаемое $g_s(x)\delta(y-y_s)$. Положив в (19) $y_1=y_s$, $y_2=y_s+\Delta y$, будем иметь при любом $\Delta y>0$

$$F_{y_s,\,y_s+\Delta y}(x)=\frac{\int\limits_{-\infty}^{x} g_s(u)\,du+O(\Delta y)}{p_s+O(\Delta y)}.$$

В пределе при $\Delta y\to 0$ получаем для условной функции распределения $F_1(x\,|\,y_s)$ и условной плотности $f_1(x\,|\,y_s)$ формулы

$$F_1(x\,|\,y_s)=\frac{1}{p_s}\int\limits_{-\infty}^{x} g_s(u)\,du,\quad f_1(x\,|\,y_s)=\frac{1}{p_s}g_s(x). \qquad (21)$$

Эти формулы справедливы и в том случае, когда $g_s(x)$ содержит линейную комбинацию δ-функций. ◀

Очевидно, что все наши рассуждения справедливы и в том случае, когда X и Y представляют собой случайные векторы. Поэтому формулы (16) и (21) определяют условное распределение проекции случайного вектора на любое подпространство при данном значении его проекции на дополнительное подпространство.

Пример 7. В условиях примера 1, подставив в (16) выражение (10) плотности $f(x, y)$ и выражение, аналогичное (12), плотности $f_2(y)$, находим условную плотность проекции двумерного нормально распределенного случайного вектора на ось x при данном значении его проекции на ось y:

$$f_1(x\,|\,y)=\sqrt{\frac{c_{11}}{2\pi}}\,e^{-\frac{1}{2}c_{11}\left[x-a+\frac{c_{12}}{c_{11}}(y-b)\right]^2}. \qquad (22)$$

Сравнив эту формулу с (3.90), видим, что условное распределение величины X при данном значении y величины Y нормально, причем математическое ожидание и дисперсия X равны соответственно

$$a-\frac{c_{12}}{c_{11}}(y-b) \quad \text{и} \quad \frac{1}{c_{11}}.$$

По симметрии заключаем, что условное распределение величины Y при данном значении x величины X нормально и при этом математическое ожидание и дисперсия Y равны соответственно

$$b-\frac{c_{12}}{c_{22}}(x-a) \quad \text{и} \quad \frac{1}{c_{22}}.$$

Пример 8. Используя результаты примера 2, найти условное распределение одной координаты n-мерного нормально распределенного случайного вектора при данных значениях всех остальных координат (т. е. при данном значении проекции случайного вектора на $(n-1)$-мерное подпространство, образованное остальными осями координат).

Пример 9. В условиях примера 4

$$f_1(x \mid y) = \frac{1}{2a\sqrt{1-y^2/b^2}} \, 1\left(a\sqrt{1-\frac{y^2}{b^2}} - |x|\right).$$

Таким образом, условное распределение величины X в данном случае равномерно в интервале $|x| < a\sqrt{1-y^2/b^2}$, длина которого зависит от y. Мы видели в примере 4, что безусловное распределение величины X в этом случае не является равномерным.

Читатель легко самостоятельно убедится в том, что и в более сложном примере 3.9 условные распределения проекций случайного вектора равномерны на соответствующих интервалах.

Пример 10. В условиях примера 5 очевидно, что условное распределение величины Y при данном значении x величины X определяется плотностью

$$f_2(y \mid x) = \delta(y - \varphi(x)).$$

Действительно, при данном x величина Y с вероятностью 1 принимает единственное возможное значение $\varphi(x)$. То же самое дает на основáнии результата примера 5 и формула (17). В частности, она справедлива в случае вырожденного распределения (2.28).

2.2. Теорема умножения плотностей. Формулы (16) и (17) можно переписать в виде

$$f(x, y) = f_1(x) f_2(y \mid x) = f_2(y) f_1(x \mid y). \tag{23}$$

Эта формула выражает теорему умножения плотностей: *совместная плотность двух случайных величин (скалярных или векторных) равна плотности одной из них, умноженной на условную плотность другой.*

► Заметим, что формула (23) верна и в том случае, когда $f_1(x)$, или $f_2(y)$, или они обе содержат δ-функции. Действительно, если, например,

$$f_2(y) = h_2(y) + \sum_{s=1}^{N} p_s \delta(y - y_s),$$

то

$$f_2(y) f_1(x \mid y) = h_2(y) f_1(x \mid y) + \sum_{s=1}^{N} p_s f_1(x \mid y_s) \delta(y - y_s). \tag{24}$$

В сумме мы заменили $f_1(x \mid y)$ функцией $f_1(x \mid y_s)$ ввиду того, что $\varphi(y)\delta(y - y_s) = \varphi(y_s)\delta(y - y_s)$ для любой функции $\varphi(y)$, поскольку $\delta(y - y_s) = 0$ при всех $y \neq y_s$. Так как при всех $y \neq y_1, \ldots, y_N$ суммы в обеих последних формулах равны нулю и, следовательно, $f_2(y) = h_2(y)$, то вследствие (16) правая часть формулы (24) равна $f(x, y)$ при всех $y \neq y_1, \ldots, y_N$. Сумма же в (24) на основании

(21) представляет собой как раз ту линейную комбинацию δ-функций, которую следует ввести в функцию $f(x, y)$, чтобы учесть отличные от нуля вероятности $p_1, \ldots, p_N$ значений $y_1, \ldots, y_N$ случайной величины Y. Таким образом, правая часть формулы (24) совпадает с $f(x, y)$ при всех x и y. ◀

2.3. Зависимые и независимые случайные величины. Случайные величины X и Y называются *независимыми*, если события $X \in A$ и $Y \in B$ независимы для любых множеств A и B, для которых определены соответствующие вероятности. Случайные величины X и Y называются *зависимыми*, если существуют множества A и B, для которых события $X \in A$ и $Y \in B$ зависимы. Эти определения относятся как к скалярным, так и к векторным случайным величинам.

Из этих определений и условий независимости событий § 1.6 следует, что случайные величины X и Y независимы тогда и только тогда, когда никакая информация об одной из них не изменяет распределение другой.

▶ Если случайные величины X и Y независимы, то согласно определению события $X < x$ и $Y < y$ независимы при любых x и y. Следовательно, в силу теоремы умножения вероятностей независимых событий (1.21)

$$P(\{X < x\}\{Y < y\}) = P(X < x)\,P(Y < y),$$

или

$$F(x, y) = F_1(x)\,F_2(y). \quad ◀ \tag{25}$$

Таким образом, *совместная функция распределения независимых случайных величин равна произведению их функций распределения.*

▶ На основании соотношений (2.32) и (2.31) между функциями распределения и плотностями из (25) следует, что

$$f(x, y) = f_1(x)\,f_2(y) \tag{26}$$

(в случае векторных X и Y для получения соотношения (26) формулу (25) следует продифференцировать по одному разу по каждой координате вектора y). ◀

Таким образом, каждое из условий (25) и (26) необходимо для независимости случайных величин X и Y.

▶ Покажем, что (26) и достаточно. Для этого вычислим вероятность пересечения событий $X \in A$ и $Y \in B$, пред-

полагая, что выполнено условие (26). Подставив выражение (26) в формулу (2.9), получаем *)

$$P(\{X\in A\}\{Y\in B\}) = P((X,\ Y)\in A\times B) = \iint\limits_{A\times B} f(x,\ y)\,dx\,dy = \int\limits_A dx\int\limits_B f_1(x)\,f_2(y)\,dy = \int\limits_A f_1(x)\,dx\int\limits_B f_2(y)\,dy,$$

или, вследствие (2.9),

$$P(\{X\in A\}\{Y\in B\}) = P(X\in A)\,P(Y\in B)$$

при любых A и B. Отсюда следует, что события $X\in A$ и $Y\in B$ независимы при любых A и B, если выполнено (26), что и доказывает достаточность условия (26) для независимости величин X и Y. Условие (25), очевидно, также достаточно, так как из (25) следует и (26). ◀

Таким образом, *каждое из условий* (25) *и* (26) *необходимо и достаточно для независимости случайных величин X и Y.*

Сравнив (26) с (22), заключаем, что *для независимости случайных величин X и Y необходимо и достаточно выполнение условий*

$$f_1(x\,|\,y) \equiv f_1(x), \qquad f_2(y\,|\,x) \equiv f_2(y). \tag{27}$$

Очевидно, что оба эти условия необходимы, но достаточно выполнения одного из них, так как в силу (22) каждое из тождеств (27) является следствием другого.

Пример 11. В примерах 1 и 7 случайные величины X и Y зависимы, если $c_{12}\neq 0$, и независимы, если $c_{12}=0$. Из (15) следует, что величины X и Y коррелированы, если $c_{12}\neq 0$, и не коррелированы, если $c_{12}=0$. Таким образом, координаты нормально распределенного двумерного случайного вектора зависимы, если они коррелированы, и независимы, если они не коррелированы.

Пример 12. В примере 3 проекции случайного вектора независимы.

Пример 13. В примерах 4 и 9 проекции случайного вектора зависимы, несмотря на то что, как мы видели в примере 3.8, они не коррелированы. В более сложном случае примера 3.9 проекции случайного вектора зависимы и в то же время коррелированы.

Пример 14. В примере 5 проекции случайного вектора X и Y зависимы. Это следует из формального сравнения $f_2(y)$ и $f_2(y\,|\,x)$, полученных в примерах 5 и 10. При этом X и Y в общем случае коррелированы. В частных случаях они могут быть и некоррелиро-

---

*) $A\times B$ представляет собой прямое произведение множеств A и B, т. е. множество упорядоченных пар $\{x,\ y\}$ таких, что $x\in A$, $y\in B$.

ванными. Так, в примере 3.10 и в более общем случае четных функций $g(x)$ и $\varphi(x)$ величины X и Y не коррелированы. В случае четной $g(x)$ и нечетной $\varphi(x)$ величины X и Y коррелированы.

Случайные величины $X_1, \ldots, X_n$ называются *независимыми*, если события $X_k \in A_k$ $(k = 1, \ldots, n)$ независимы для любых множеств $A_1, \ldots, A_n$, для которых определены соответствующие вероятности. Случайные величины $X_1, \ldots, X_n$ называются *зависимыми*, если существуют множества $A_1, \ldots, A_n$, для которых события $X_k \in A_k$ $(k = 1, \ldots, n)$ зависимы. Эти определения относятся как к скалярным, так и к векторным величинам $X_1, \ldots, X_n$.

Так же как в случае двух величин, из этого определения следует, что для независимости $X_1, \ldots, X_n$ необходимо и достаточно, чтобы их совместная функция распределения (плотность, если она существует) была равна произведению их функций распределения (соответственно плотностей).

Попарной независимости величин $X_1, \ldots, X_n$ недостаточно для того, чтобы они были независимыми.

Пример 15. Пусть X_1 и X_2 — независимые случайные величины, каждая из которых имеет два возможных значения 0 и 1 с вероятностями 1/2, $X_3 = X_1 X_2 + (1 - X_1)(1 - X_2)$. Величина X_3 также имеет два возможных значения 0 и 1, вероятности которых равны 1/2, так как $P(X_3 = 0) = P(X_1 = 0) P(X_2 = 1) + P(X_1 = 1) P(X_2 = 0) = 1/2$, $P(X_3 = 1) = 1 - P(X_3 = 0) = 1/2$. Величины X_1, X_2 и X_3 попарно независимы, так как

$$P(X_1=0,\ X_3=0) = P(X_1=0) P(X_2=1) = \frac{1}{4} = P(X_1=0) P(X_3=0),$$

$$P(X_1=0,\ X_3=1) = P(X_1=0) P(X_2=0) = \frac{1}{4} = P(X_1=0) P(X_3=1),$$

$$P(X_1=1,\ X_3=0) = P(X_1=1) P(X_2=0) = \frac{1}{4} = P(X_1=1) P(X_3=0),$$

$$P(X_1=1,\ X_3=1) = P(X_1=1) P(X_2=1) = \frac{1}{4} = P(X_1=1) P(X_3=1),$$

и аналогичные равенства справедливы для X_2, X_3. Однако

$$P(X_1 = 0,\ X_2 = 0,\ X_3 = 0) = 0 \neq P(X_1 = 0) P(X_2 = 0) P(X_3 = 0).$$

Следовательно, величины X_1, X_2 и X_3 зависимы.

Пример 16. Трехмерный случайный вектор $X = \{X_1, X_2, X_3\}$, равномерно распределенный на поверхности тетраэдра с вершинами в точках (0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0), имеет плотность

$$f(x) = \frac{1}{2} 1_{(0,1)}(x_1) 1_{(0,1)}(x_2) 1_{(0,1)}(x_3) [\delta(x_1 + x_2 - x_3) + \\ + \delta(x_1 + x_3 - x_2) + \delta(x_2 + x_3 - x_1) + \delta(x_1 + x_2 + x_3 - 2)],$$

где $1_{(0,1)}(x)$ — индикатор интервала $(0, 1)$. Любая координата этого вектора распределена равномерно в интервале $(0, 1)$, а его проекция на любую координатную плоскость распределена равномерно на квадрате $(0, 1)\times(0, 1)=(0, 1)^2$. Следовательно, величины X_1, X_2 и X_3 попарно независимы. В то же время они зависимы, так как в противном случае распределение вектора X было бы равномерным в кубе $(0, 1)^3$.

Пример 17. Плотность n-мерного случайного вектора X определяется формулой

$$f(x)=f_1(x_1)\dots f_n(x_n)+g(x),$$

где $f_1(x), \dots, f_n(x)$ — любые плотности (возможно, содержащие линейные комбинации δ-функций), а $g(x)$ — любая функция, удовлетворяющая условиям:

1) $|g(x)|\leqslant f_1(x_1)\dots f_n(x_n)$ при всех $x_1, \dots, x_n$;

2) $\int\limits_{-\infty}^{\infty} g(x)\,dx_k=0$ при всех x_i, $i\neq k$ $(k=1, \dots, n)$.

Этим условиям удовлетворяет, например, функция

$$g(x)=g_1(x_1)\dots g_n(x_n),$$

где $g_1(x), \dots, g_n(x)$ — любые нечетные функции, модули которых при всех x меньше соответствующих функций $f_1(x), \dots, f_n(x)$. Интегрируя плотность $f(x)$ по какой-нибудь одной координате, убеждаемся в том, что любые $n-1$ из координат вектора X независимы, причем $f_k(x_k)$ есть плотность координаты X_k $(k=1, \dots, n)$. Однако все n координат вектора X зависимы, если $g(x)\not\equiv 0$.

Подобных примеров можно привести много. В частности, этот пример легко распространяется на векторные величины $X_1, \dots, X_n$.

2.4. Независимые величины не коррелированы. Докажем, что *независимые случайные величины всегда не коррелированы.*

▶ Подставив выражение (26) совместной плотности независимых случайных величин X и Y в формулу (3.22) для ковариации, получаем

$$k_{xy}=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(x-m_x)(y-m_y)f_1(x)f_2(y)\,dx\,dy=$$

$$=\int\limits_{-\infty}^{\infty}(x-m_x)f_1(x)\,dx\int\limits_{-\infty}^{\infty}(y-m_y)f_2(y)\,dy.$$

Но

$$\int\limits_{-\infty}^{\infty}(y-m_y)f_2(y)\,dy=\int\limits_{-\infty}^{\infty}yf_2(y)\,dy-m_y\int\limits_{-\infty}^{\infty}f_2(y)\,dy=0.$$

Следовательно, $k_{xy}=0$, что и доказывает наше утверждение. ◀

Зависимые случайные величины, как показывают примеры 11—14, могут быть коррелированными, а могут быть и некоррелированными.

2.5. Независимость функций независимых величин. Докажем, что *если величины $X_1, \ldots, X_n$ независимы, то и любые их функции $Y_1=\varphi_1(X_1), \ldots, Y_n=\varphi_n(X_n)$ независимы.*

▶ Действительно, в этом случае

$$P\left(\bigcap_{k=1}^{n}\{\varphi_k(X_k)\in B_k\}\right)=$$
$$=\int\limits_{\varphi_1(x_1)\in B_1}\cdots\int\limits_{\varphi_n(x_n)\in B_n} f_1(x_1)\ldots f_n(x_n)\,dx_1\ldots dx_n=$$
$$=\prod_{k=1}^{n}\int\limits_{\varphi_k(x_k)\in B_k} f_k(x_k)\,dx_k=\prod_{k=1}^{n}P(\varphi_k(X_k)\in B_k)$$

для любых множеств $B_1, \ldots, B_n$ в соответствующих пространствах. ◀

2.6. Теорема умножения математических ожиданий. В п. 3.2.4 была выведена формула (3.30), связывающая математические ожидания, ковариацию и смешанный начальный момент второго порядка двух случайных величин. В случае действительных величин X и Y эта формула имеет вид

$$\gamma_{xy}=k_{xy}+m_x m_y.$$

Отсюда следует, что если случайные величины X и Y не коррелированы ($k_{xy}=0$), то $\gamma_{xy}=m_x m_y$, или

$$MXY=MX\cdot MY. \qquad (28)$$

Таким образом, *математическое ожидание произведения двух некоррелированных действительных случайных величин равно произведению их математических ожиданий.* Эта теорема умножения математических ожиданий справедлива, в частности, для независимых случайных величин, так как независимые величины, как было показано в п. 2.4, всегда не коррелированы.

Естественно возникает желание обобщить эту теорему на любое число величин. Однако некоррелированности и даже попарной независимости случайных величин для этого уже недостаточно.

Пример 18. В условиях примера 15 величины X_1, X_2 и X_3 попарно независимы, а следовательно, и не коррелированы. Их произведение имеет два возможных значения 0 и 1, причем значение 1 достигается только при $X_1=X_2=1$, а вероятность этого равна 1/4. Следовательно, $MX_1X_2X_3=1/4$. В то же время $MX_1=MX_2=MX_3=1/2$ и $MX_1\cdot MX_2\cdot MX_3=1/8$. Следовательно, теорема умножения математических ожиданий в данном случае не справедлива.

▶ Если случайные величины $X_1, \ldots, X_n$ независимы, то для любых функций $\varphi_1(X_1), \ldots, \varphi_n(X_n)$

$$M\varphi_1(X_1)\varphi_2(X_2)\ldots\varphi_n(X_n)=$$
$$=\int\limits_{-\infty}^{\infty}\ldots\int\limits_{-\infty}^{\infty}\varphi_1(x_1)\varphi_2(x_2)\ldots$$
$$\ldots\varphi_n(x_n)f_1(x_1)f_2(x_2)\ldots f_n(x_n)\,dx_1\ldots dx_n=$$
$$=\int\limits_{-\infty}^{\infty}\varphi_1(x_1)f_1(x_1)\,dx_1\int\limits_{-\infty}^{\infty}\varphi_2(x_2)f_2(x_2)\,dx_2\ldots$$
$$\ldots\int\limits_{-\infty}^{\infty}\varphi_n(x_n)f(x_n)\,dx_n=M\varphi_1(X_1)\,M\varphi_2(X_2)\ldots M\varphi_n(X_n),$$

причем математическое ожидание в левой части существует тогда и только тогда, когда существуют все математические ожидания в правой части. ◀

Таким образом, *если случайные величины $X_1, \ldots, X_n$ независимы, то математическое ожидание произведения любых n функций этих величин, каждая из которых зависит только от одной из величин $X_1, \ldots, X_n$, равно произведению математических ожиданий этих функций*:

$$M\varphi_1(X_1)\varphi_2(X_2)\ldots\varphi_n(X_n)=$$
$$=M\varphi_1(X_1)\,M\varphi_2(X_2)\ldots M\varphi_n(X_n). \quad (29)$$

Эта теорема справедлива как для скалярных, так и для векторных величин $X_1, \ldots, X_n$ и для произвольных комплекснозначных функций $\varphi_1, \ldots, \varphi_n$.

В частном случае действительных величин $X_1, \ldots, X_n$ можно взять $\varphi_1(X_1)=X_1, \ldots, \varphi_n(X_n)=X_n$, и мы получаем обобщение теоремы умножения математических ожиданий на любое число случайных величин: *математическое ожидание произведения независимых действительных случайных величин равно произведению их математических ожиданий.*

Так как комплексную случайную величину можно рассматривать как функцию двумерного случайного век-

тора, координатами которого служат ее действительная и мнимая части, то из (29) следует, что теорема умножения математических ожиданий справедлива и для независимых комплексных случайных величин.

Пример 19. Показать, что теорема умножения математических ожиданий справедлива для действительных случайных величин $X_1, \ldots$ $\ldots, X_n$, если все их смешанные центральные моменты до порядка n включительно равны 0.

§ 3. Условные числовые характеристики

3.1. Условное математическое ожидание. Все введенные в главе 3 числовые характеристики можно определить и для условных распределений.

В частности, из формулы (3.3) вытекает определение *условного математического* ожидания данной функции $\varphi(Y)$ случайной величины Y *при данном значении* x *случайной величины* X:

$$M[\varphi(Y) \mid x] = \int_{-\infty}^{\infty} \varphi(y) f_2(y \mid x)\, dy, \tag{30}$$

где $f_2(y \mid x)$ — условная плотность величины Y при данном значении x величины X. Эта формула определяет условное математическое ожидание как в случае скалярных, так и в случае векторных величин X, Y и $Z = \varphi(Y)$.

Из формулы (30) как частный случай вытекает формула для условного математического ожидания случайной величины Y при данном значении x величины X:

$$M[Y \mid x] = \int_{-\infty}^{\infty} y f_2(y \mid x)\, dy. \tag{31}$$

3.2. Регрессия. Обозначение $f_2(y \mid x)$ применяется для плотности случайной величины Y, зависящей от параметра x, и в том случае, когда x не является значением некоторой случайной величины X. Формулы (30) и (31) в этом случае определяют математические ожидания величин Y и $\varphi(Y)$ как функции параметра x.

Математическое ожидание случайной величины Y как функция параметра x, от которого зависит распределение Y, называется *регрессией* Y *на* x. В частном случае, когда параметр x представляет собой возможное значение некоторой случайной величины X, регрессия Y на x пред-

ставляет собой условное математическое ожидание величины Y при $X=x$.

График функции $y=M[Y|x]$ в случае скалярных x и Y называется *кривой регрессии Y на x*.

Если параметр x представляет собой значение случайной величины X, то, кроме регрессии Y на x, можно определить также регрессию X на y.

Определив условное математическое ожидание, можно определить и все *условные моменты* случайных величин. Ясно, что все условные характеристики случайных величин при данном значении x случайной величины X в общем случае являются определенными функциями x. Поэтому их можно рассматривать до опыта как функции случайной величины X. В этом случае они сами будут случайными величинами.

Условное математическое ожидание случайной величины $Z=\varphi(Y)$, рассматриваемое как функция случайной величины X, $M[\varphi(Y)|X]$, называется *условным математическим ожиданием* случайной величины $Z=\varphi(Y)$ *относительно* X.

Так как условные математические ожидания и другие условные моменты случайных величин относительно случайной величины X сами являются случайными величинами, то для них, в свою очередь, можно определить числовые характеристики, в частности математические ожидания и моменты.

Пример 20. В примерах 3.9, 4 и 9

$$f_2(y|x)=\frac{1}{2b\sqrt{1-x^2/a^2}} \quad \text{при} \quad |y-cx|<\sqrt{1-\frac{x^2}{a^2}}$$

и $f_2(y|x)=0$ при всех других значениях y. Поэтому

$$M[Y|x]=\frac{1}{2b\sqrt{1-x^2/a^2}}\int\limits_{cx-b\sqrt{1-x^2/a^2}}^{cx+b\sqrt{1-x^2/a^2}} y\,dy=cx.$$

Таким образом, условное математическое ожидание Y при $X=x$ в данном случае равно cx.

Условная дисперсия величины Y при $X=x$, в соответствии с результатом примера 3.3, определяется формулой

$$D[Y|x]=\frac{b^2}{3a^2}(a^2-x^2).$$

Предлагаем читателю самостоятельно доказать, что в этом случае

$$M[X|y]=\frac{a^2c^2}{b^2+a^2c^2}y.$$

Пример 21. В случае двумерного нормального распределения (примеры 1 и 7) условное математическое ожидание и условная дисперсия величины Y при $X=x$ определяются формулами

$$M[Y \mid x] = b - \frac{c_{12}}{c_{22}}(x-a), \qquad D[Y \mid x] = \frac{1}{c_{22}}.$$

Таким образом, уравнение кривой регрессии Y на x имеет в данном случае вид

$$y = b - \frac{c_{12}}{c_{22}}(x-a).$$

Аналогично находим уравнение кривой регрессии X на y:

$$x = a - \frac{c_{12}}{c_{11}}(y-b).$$

Легко видеть, что кривая регрессии Y на x представляет собой диаметр эллипса рассеивания, сопряженный с направлением оси y, а кривая регрессии X на y — диаметр, сопряженный с направлением оси x.

3.3. Формула полного математического ожидания. Выведем одну часто применяемую в теории вероятностей формулу.

► Применив формулу (3.3) для вычисления математического ожидания функции $\varphi(X, Y)$ случайного вектора $\{X, Y\}$, можем написать

$$M\varphi(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x, y) f(x, y)\, dx\, dy, \tag{32}$$

где $f(x, y)$ — плотность случайного вектора $\{X, Y\}$. Подставив сюда выражение $f(x, y)$ из (23):

$$f(x, y) = f_1(x) f_2(y \mid x),$$

будем иметь

$$M\varphi(X, Y) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \varphi(x, y) f_2(y \mid x)\, dy \right\} f_1(x)\, dx.$$

Но интеграл в фигурных скобках представляет собой условное математическое ожидание случайной величины $\varphi(x, Y)$ при данном значении x величины X:

$$\int_{-\infty}^{\infty} \varphi(x, y) f_2(y \mid x)\, dy = M[\varphi(x, Y) \mid x].$$

Следовательно, предыдущую формулу можно переписать в виде

$$M\varphi(X, Y) = \int_{-\infty}^{\infty} M[\varphi(x, Y) | x] f_1(x)\, dx. \qquad (33)$$

Правая часть этого равенства представляет собой математическое ожидание функции

$$\psi(X) = M[\varphi(X, Y) | X]$$

случайной величины X. Следовательно,

$$M\varphi(X, Y) = M[M[\varphi(X, Y) | X]]. \blacktriangleleft \qquad (34)$$

Эта формула называется *формулой полного математического ожидания*. Она показывает, что при вычислении математического ожидания функции двух случайных величин можно сначала найти условное математическое ожидание этой функции при фиксированном значении одной из величин-аргументов, а потом найти математическое ожидание этого условного математического ожидания, рассматриваемого как функция этой случайной величины.

Формула (34) справедлива как для скалярных, так и для векторных случайных величин X и Y. В последнем случае все интегралы в предыдущих формулах следует понимать как кратные интегралы по соответствующим пространствам.

Формула (34) верна и для комплексных случайных величин (в том числе векторных) X и Y. Чтобы убедиться в этом, достаточно понимать $f_1(x)$ и $f_2(y | x)$ в предыдущих выкладках как плотности случайных векторов, образованных действительными и мнимыми частями величин X и Y (или всех их координат).

В частном случае при $\varphi(X, Y) = Y$ формула (34) дает

$$MY = M[M[Y | X]]. \qquad (35)$$

Таким образом, *математическое ожидание условного математического ожидания случайной величины Y относительно X равно безусловному математическому ожиданию величины Y*.

В частном случае дискретной величины X, подставив в (33) выражение (2.24) плотности $f_1(x)$, получим

$$M\varphi(X, Y) = \sum_{k=1}^{N} p_k M[Y | x_k]. \qquad (36)$$

П р и м е р 22. Число приборов, поступающих в течение дня в ремонтную мастерскую, представляет собой случайную величину, распределенную по закону Пуассона с математическим ожиданием μ. Вероятность того, что поступивший прибор потребует большого ремонта, равна p. Найти математическое ожидание числа поступающих за день приборов, требующих большого ремонта.

При данном числе n поступивших за день приборов число приборов, требующих большого ремонта, представляет собой случайную величину, распределенную по биномиальному закону. Поэтому, в соответствии с результатом примера 3.1,

$$M[Y \mid n] = np \qquad (n = 0, 1, 2, \ldots).$$

Вероятность того, что в течение дня поступит n приборов, в соответствии с (1.46), определяется формулой

$$p_n = \frac{\mu^n}{n!} e^{-\mu} \qquad (n = 0, 1, 2, \ldots).$$

Подставив полученные выражения вероятностей и условных математических ожиданий в формулу (36), получим

$$M[Y] = \sum_{n=0}^{\infty} p_n M[Y \mid n] = \sum_{n=1}^{\infty} \frac{\mu^n}{n!} e^{-\mu} np =$$

$$= \mu p e^{-\mu} \sum_{n=1}^{\infty} \frac{\mu^{n-1}}{(n-1)!} = \mu p e^{-\mu} \cdot e^{\mu} = \mu p.$$

§ 4. Многомерное нормальное распределение

4.1. Математическое ожидание нормально распределенного вектора. Изучим нормально распределенные случайные векторы. В соответствии с (2.22) напишем выражение плотности n-мерного нормально распределенного вектора X в виде

$$f(x) = A_n e^{-\frac{1}{2}(x^T - a^T)C(x-a)}. \tag{37}$$

Результаты изучения одномерного и двумерного нормальных распределений в § 3.6 и в примерах 1 и 7 наводят на мысль, что параметр a в (37) представляет собой математическое ожидание случайного вектора X, $a = m_x = MX$, матрица коэффициентов C квадратичной формы в показателе степени обратна по отношению к ковариационной матрице K_x вектора X, а коэффициент A_n при показательной функции определяется формулой

$$A_n = \sqrt{\frac{|C|}{(2\pi)^n}} = \frac{1}{\sqrt{(2\pi)^n |K_x|}}. \tag{38}$$

Так как это верно для $n=1$ и $n=2$, то для доказательства этого утверждения в общем случае проще всего применить метод индукции. Предположив, что утверждение верно для $(n-1)$-мерного вектора, докажем, что в таком случае оно верно и для n-мерного вектора.

▶ Пусть X' — случайный вектор, образованный первыми $n-1$ координатами вектора X, $X' = \{X_1, \ldots, X_{n-1}\}$, и представим плотность $f(x)$ в соответствии с (23) в виде произведения плотности $f_1(x')$ вектора X' и условной плотности последней координаты X_n вектора X:

$$f(x) = f_1(x')\, f_2(x_n \mid x'). \tag{39}$$

Для этого достаточно разложить $f(x)$ на два множителя, один из которых не зависит от x_n, а другой при интегрировании по x_n в бесконечных пределах дает постоянную величину, не зависящую от x'. Тогда выбором подходящих числовых множителей к этим двум функциям получим требуемый результат.

Выделим в показателе степени в (37) все слагаемые, зависящие от x_n. Для этого положим для краткости $u = x - a$ и представим матрицу коэффициентов C в виде блочной матрицы:

$$C = \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1,\,n-1} & c_{1n} \\ c_{12} & c_{22} & \ldots & c_{2,\,n-1} & c_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{1,\,n-1} & c_{2,\,n-1} & \ldots & c_{n-1,\,n-1} & c_{n-1,\,n} \\ c_{1n} & c_{2n} & \ldots & c_{n-1,\,n} & c_{nn} \end{bmatrix} = \begin{bmatrix} C_1 & c \\ c^T & c_{nn} \end{bmatrix},$$

где C_1 — матрица коэффициентов при $u_p u_q$ $(p, q = 1, \ldots, n-1)$, а c — матрица-столбец коэффициентов при $u_p u_n$ $(p = 1, \ldots, n-1)$. Точно так же матрицу-столбец u разобьем на два блока u' и u_n, $u^T = [u'^T \ u_n]$. Тогда будем иметь

$$u^T C u = [u'^T \ u_n] \begin{bmatrix} C_1 & c \\ c^T & c_{nn} \end{bmatrix} \begin{bmatrix} u' \\ u_n \end{bmatrix} =$$
$$= [u'^T \ u_n] \begin{bmatrix} C_1 u' + c u_n \\ c^T u' + c_{nn} u_n \end{bmatrix} = u'^T C_1 u' + u'^T c u_n +$$
$$+ u_n c^T u' + c_{nn} u_n^2 = u'^T C_1 u' + 2 u_n c^T u' + c_{nn} u_n^2,$$

так как $u'^T c = c^T u'$ *). Дополнив два последних слагаемых до полного квадрата и снова пользуясь равенством

---

*) Читатель, которого затрудняют выкладки с блочными матрицами, легко убедится в справедливости этого равенства, представив его в скалярной форме.

$c^T u' = u'^T c$, получим

$$u^T C u = u'^T C_1 u' + c_{nn}\left[u_n^2 + 2u_n \frac{c^T u'}{c_{nn}} + \left(\frac{c^T u'}{c_{nn}}\right)^2\right] -$$
$$- \frac{1}{c_{nn}} (c^T u')^2 = u'^T C_1 u' - u'^T \frac{cc^T}{c_{nn}} u' + c_{nn}\left(u_n + \frac{c^T u'}{c_{nn}}\right)^2.$$

Вводя для краткости матрицу

$$C' = C_1 - \frac{cc^T}{c_{nn}}, \tag{40}$$

перепишем полученное равенство в виде

$$u^T C u = u'^T C' u' + c_{nn}\left(u_n + \frac{c^T u'}{c_{nn}}\right)^2. \tag{41}$$

Подставив это выражение в (37) и вспомнив, что $u = x - a$ и, соответственно, $u' = x' - a'$, получим

$$f(x) = A_n \exp\left\{-\frac{1}{2}(x'^T - a'^T) C' (x' - a')\right\} \times$$
$$\times \exp\left\{-\frac{1}{2} c_{nn}\left[x_n - a_n + \frac{c^T}{c_{nn}}(x' - a')\right]^2\right\}. \tag{42}$$

Интеграл от второй показательной функции по x_n в бесконечных пределах не зависит от x', в чем можно убедиться, приняв за переменную интегрирования выражение в квадратных скобках вместо x_n. Следовательно, эта показательная функция пропорциональна условной плотности $f_2(x_n | x')$ случайной величины X_n при данном значении x' величины X'. Сравнив эту показательную функцию с выражением (3.85) одномерной нормальной плотности, получаем

$$f_2(x_n | x') = \sqrt{\frac{c_{nn}}{2\pi}} \exp\left\{-\frac{1}{2} c_{nn}\left[x_n - a_n + \frac{c^T}{c_{nn}}(x' - a')\right]^2\right\}. \tag{43}$$

Оставшийся множитель в (42) представляет собой плотность случайного вектора X':

$$f_1(x') = A_n \sqrt{\frac{2\pi}{c_{nn}}} \exp\left\{-\frac{1}{2}(x'^T - a'^T) C' (x' - a')\right\}. \tag{44}$$

Из (41) следует, что матрица C' положительно определенна, так как при любом векторе u' и $u_n = -(c^T u')/c_{nn}$ имеем $u'^T C' u' = u^T C u > 0$. Следовательно, распределение вектора X' нормально. А так как по предположению для $(n-1)$-мерного нормального распределения наше

утверждение справедливо, то

$$A_{n-1}=A_n\sqrt{\frac{2\pi}{c_{nn}}}=\frac{1}{\sqrt{(2\pi)^{n-1}|K_{x'}|}}, \tag{45}$$

$$C'K_{x'}=I, \tag{46}$$

где $K_{x'}$ — ковариационная матрица случайного вектора X', и $a'=MX'$ или, что то же, $a_p=m_p=MX_p$ $(p=1, \ldots, n-1)$.

Таким образом, величины $a_1, \ldots, a_{n-1}$ равны математическим ожиданиям $m_1, \ldots, m_{n-1}$ случайных величин $X_1, \ldots, X_{n-1}$. Вследствие симметрии выражения плотности (37) относительно переменных $x_1, \ldots, x_n$ заключаем, что и величина a_n равна математическому ожиданию m_n случайной величины X_n. Поэтому $a=MX$, что и доказывает первую часть утверждения для n-мерного распределения. ◀

4.2. Ковариационная матрица. Докажем теперь, что ковариационная матрица K_x вектора X равна C^{-1}.

▶ Положим

$$Y_n=X_n-a_n+\frac{c^T}{c_{nn}}(X'-a'). \tag{47}$$

При данном значении x' вектора X' условное распределение величины Y_n, очевидно, получается соответствующим сдвигом условного распределения величины X_n. Поэтому из формулы (43) следует, что условная плотность величины Y_n при $X'=x'$ определяется формулой

$$g(y_n\,|\,x')=\sqrt{\frac{c_{nn}}{2\pi}}\,e^{-\frac{1}{2}c_{nn}y_n^2}$$

и, следовательно, не зависит от x'. Это значит, что случайная величина Y_n не зависит от случайных величин $X_1, \ldots, X_{n-1}$. Но независимые величины всегда не коррелированы. Следовательно, ковариации случайной величины Y_n со всеми величинами $X_1, \ldots, X_{n-1}$ равны нулю. Для вычисления этих ковариаций, принимая во внимание, что $a_n=MX_n$, $a'=MX'$, перепишем (47) в виде

$$Y_n=X_n^0-\sum_{p=1}^{n-1}\frac{c_{pn}}{c_{nn}}X_p^0.$$

Умножив это равенство на X_r^0 и взяв математическое ожидание, найдем ковариацию Y_n и X_r. Учитывая, что по доказанному она равна 0 при $r=1, \ldots, n-1$, получим

соотношения

$$k_{rn}+\sum_{p=1}^{n-1}\frac{c_{pn}}{c_{nn}}k_{rp}=0 \qquad (r=1,\ \ldots,\ n-1), \qquad (48)$$

откуда находим

$$\sum_{p=1}^{n}k_{rp}c_{pn}=0 \qquad (r=1,\ \ldots,\ n-1). \qquad (49)$$

Перепишем теперь (46) с учетом (40) в скалярной форме:

$$\sum_{p=1}^{n-1}k_{rp}\left(c_{ps}-\frac{c_{pn}c_{sn}}{c_{nn}}\right)=\delta_{rs} \qquad (r,\ s=1,\ \ldots,\ n-1),$$

или, так как $c_{sn}=c_{ns}$,

$$\sum_{p=1}^{n-1}k_{rp}c_{ps}-c_{ns}\sum_{p=1}^{n-1}\frac{c_{pn}}{c_{nn}}k_{rp}=\delta_{rs}.$$

Но из (48) следует, что

$$-\sum_{p=1}^{n-1}\frac{c_{pn}}{c_{nn}}k_{rp}=k_{rn}.$$

Поэтому предыдущую формулу можно переписать в виде

$$\sum_{p=1}^{n}k_{rp}c_{ps}=\delta_{rs} \qquad (r,\ s=1,\ \ldots,\ n-1). \qquad (50)$$

Формулы (49) и (50) показывают, что равенства (50) справедливы при $s=1,\ \ldots,\ n$; $r=1,\ \ldots,\ n-1$. Вследствие симметрии выражения (37) плотности относительно переменных $x_1,\ \ldots,\ x_n$, наряду с (50) справедливы и равенства, получаемые из (50) круговой подстановкой индексов $1,\ \ldots,\ n$. Следовательно, равенства (50) справедливы при всех $r,\ s=1,\ \ldots,\ n$. Это и доказывает, что матрицы C и K_x являются взаимно обратными. ◀

4.3. **Коэффициент при показательной функции.** Докажем последнюю часть утверждения — равенство (38).

▶ Заметим, что так как c_{nn} — элемент матрицы C, обратной по отношению к K_x, а алгебраическое дополнение элемента k_{nn} в определителе $|K_x|$ есть $|K_{x'}|$, то $c_{nn}=|K_{x'}|/|K_x|$. Подставив это выражение в (45) и учтя, что определители взаимно обратных матриц являются

взаимно обратными величинами, мы и получим формулу (38) для коэффициента A_n в (37). ◀

Таким образом, при любом n коэффициент A_n в выражении нормальной плотности определяется формулой (38), $a = m_x$, $C = K_x^{-1}$. Поэтому выражение (37) плотности нормально распределенного случайного вектора X можно написать в виде

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |K_x|}} \exp\left\{-\frac{1}{2}(x^T - m_x^T) K_x^{-1}(x - m_x)\right\}. \quad (51)$$

Эта формула показывает, что многомерное нормальное распределение полностью определяется математическим ожиданием и ковариационной матрицей случайного вектора. Таким образом, *нормальное распределение всегда полностью определяется моментами первого и второго порядков*.

Нормальное распределение случайного вектора с математическим ожиданием m и ковариационной матрицей K коротко обозначается $N(m, K)$. Таким образом, утверждение, что случайный вектор X имеет распределение $N(m, K)$, означает, что он распределен нормально и его математическое ожидание и ковариационная матрица равны соответственно m и K. В частности, утверждение, что скалярная (действительная) случайная величина X имеет распределение $N(m, D)$, означает, что она распределена нормально и ее математическое ожидание и дисперсия равны соответственно m и D.

Заметим, что, доказав формулу (38), мы попутно вычислили сложный n-кратный интеграл

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^T C u}\, du = \sqrt{(2\pi)^n |K_x|} = \sqrt{\frac{(2\pi)^n}{|C|}}. \quad (52)$$

Этот результат является одним из многочисленных примеров решения математических задач чисто вероятностными методами. Используя вероятностные закономерности, часто удается вычислять сложные интегралы без выполнения интегрирований или упрощать их до такой степени, чтобы дальнейшее интегрирование стало элементарным, а также решать многие другие задачи.

4.4. Условные распределения координат. Формула (43) показывает, что условное распределение каждой координаты нормально распределенного случайного вектора при данных значениях всех остальных координат нор-

мально, причем условное математическое ожидание и условная дисперсия координаты X_n определяются формулами

$$M[X_n \mid x'] = a_n - \frac{c^T}{c_{nn}}(x' - a') = m_n - \sum_{p=1}^{n-1} \frac{c_{pn}}{c_{nn}}(x_p - m_p), \tag{53}$$

$$D[X_n \mid x'] = \frac{1}{c_{nn}}. \tag{54}$$

Таким образом, от x' зависит только условное математическое ожидание величины X_n, представляющее собой сдвинутую линейную функцию x', т. е. переменных $x_1, \ldots, x_{n-1}$.

Из установленных фактов как следствие вытекает, что *распределения и условные распределения проекций нормально распределенного случайного вектора на любые подпространства нормальны.*

▶ Переписав формулу (53) в виде

$$M[X_n \mid x'] = m_n + g^T(x' - m') = m_n + \sum_{p=1}^{n-1} g_p(x_p - m_p), \tag{55}$$

на основании соотношений (50) при $s = n$, $r = 1, \ldots, n-1$, убеждаемся в том, что величины $g_p = -c_{pn}/c_{nn}$ удовлетворяют уравнениям

$$\sum_{p=1}^{n-1} k_{rp} g_p = k_{rn} \qquad (r = 1, \ldots, n-1), \tag{56}$$

или в векторной форме

$$K_{x'} g = k, \tag{57}$$

где k — вектор с координатами $k_{1n}, \ldots, k_{n-1,n}$. ◀

Таким образом, *условное математическое ожидание случайной величины X_n при $X' = x'$ представляет собой сдвинутую линейную функцию x', коэффициенты которой определяются системой линейных алгебраических уравнений* (56) *или* (57).

4.5. Случай некоррелированных координат. Рассмотрим частный случай, когда координаты случайного вектора X не коррелированы.

▶ В этом случае $k_{rp} = 0$ при $r \neq p$ и, следовательно,

$$c_{pq} = 0 \text{ при } q \neq p, \quad c_{pp} = \frac{1}{k_{pp}} \quad (p = 1, \ldots, n).$$

Выражение (37) плотности при этом принимает вид

$$f(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n k_{11} \ldots k_{nn}}} e^{-\frac{(x_1-a_1)^2}{2k_{11}}} \ldots e^{-\frac{(x_n-a_n)^2}{2k_{nn}}} =$$
$$= f_1(x_1) f_2(x_2) \ldots f_n(x_n).$$

Отсюда видно, что в случае некоррелированных координат нормально распределенного случайного вектора его плотность равна произведению плотностей его координат. Следовательно, координаты случайного вектора в этом случае независимы. ◀

Таким образом, *координаты нормально распределенного случайного вектора не коррелированы тогда и только тогда, когда они независимы.*

4.6. Вырожденное нормальное распределение. Формулы (37) и (51) определяют плотность для нормального распределения только в том случае, когда определитель матрицы K_x не равен нулю и, следовательно, ранг матрицы K_x совпадает с размерностью n случайного вектора X. Если ранг матрицы K_x равен $r < n$, то, как мы знаем из результатов § 3.3, между координатами вектора X^0 существует $n-r$ линейных зависимостей, вследствие чего распределение полностью сосредоточено на r-мерном подпространстве, сдвинутом на вектор m_x (если m_x не принадлежит этому r-мерному подпространству). Выражение плотности случайного вектора X в этом случае, как мы видели в п. 2.3.4, содержит множителем δ-функцию $(n-r)$-мерного аргумента.

▶ Если ранг матрицы K_x равен $r < n$, то существует отличный от нуля диагональный определитель порядка r матрицы K_x. Соответствующая обратимая матрица $K_{x'}$ определяет нормальное распределение проекции X' случайного вектора на подпространство, образованное соответствующими осями координат:

$$f_1(x') = \frac{1}{\sqrt{(2\pi)^r |K_{x'}|}} \exp\left\{-\frac{1}{2}(x'-m')^T K_{x'}^{-1}(x'-m')\right\}, \tag{58}$$

где $m' = MX'$. Проекция X'' случайного вектора X на дополнительное $(n-r)$-мерное подпространство однозначно определяется из уравнений $\varphi_p^T X^0 = 0$ $(p = r+1, \ldots, n)$, где $\varphi_{r+1}, \ldots, \varphi_n$ — собственные векторы матрицы K_x, соответствующие $(n-r)$-кратному нулевому собственному значению. Чтобы решить эти уравнения, представим

матрицы-столбцы φ_p, X и m_x в блочном виде:

$$\varphi_p^T = [\varphi_p'^T \ \ \varphi_p''^T], \quad X^T = [X'^T \ \ X''^T], \quad m_x^T = [m'^T \ \ m''^T].$$

Тогда уравнения примут вид

$$\varphi_p'^T X'^0 + \varphi_p''^T X''^0 = 0 \qquad (p = r+1, \ldots, n). \tag{59}$$

Собственные векторы $\varphi_{r+1}, \ldots, \varphi_n$ определяются уравнениями $K_x \varphi_p = 0$ $(p = r+1, \ldots, n)$. Представив матрицу K_x в блочном виде:

$$K_x = \begin{bmatrix} K_{x'} & K_{x'x''} \\ K_{x''x'} & K_{x''} \end{bmatrix},$$

можем написать эти уравнения в виде

$$K_{x'}\varphi_p' + K_{x'x''}\varphi_p'' = 0, \quad K_{x''x'}\varphi_p' + K_{x''}\varphi_p'' = 0 \ (p = r+1, \ldots, n). \tag{60}$$

Так как матрица $K_{x'}$ обратима, первое уравнение можно решить относительно φ_p'. В результате получим $\varphi_p' = -K_{x'}^{-1} K_{x'x''} \varphi_p''$. Отсюда, принимая во внимание, что в силу (3.42) $K_{x'x''}^T = K_{x''x'}$, получаем $\varphi_p'^T = -\varphi_p''^T K_{x''x'} K_{x'}^{-1}$. Подставив это выражение в (59), получаем

$$\varphi_p''^T (X''^0 - K_{x''x'} K_{x'}^{-1} X'^0) = 0 \qquad (p = r+1, \ldots, n). \tag{61}$$

Таким образом, мы получили систему однородных линейных уравнений относительно координат вектора $X''^0 - K_{x''x'} K_{x'}^{-1} X'^0$. Докажем, что определитель этой системы не равен нулю. Для этого подставим выражение $\varphi_p'^T$ из первого уравнения (60) в соотношение ортогональности векторов $\{\varphi_p\}$:

$$\varphi_p'^T \varphi_q' + \varphi_p''^T \varphi_q'' = \delta_{pq}.$$

Получим

$$\varphi_p''^T (\varphi_q'' - K_{x''x'} K_{x'}^{-1} \varphi_q') = \delta_{pq} \qquad (p = r+1, \ldots, n).$$

При любом фиксированном $q = r+1, \ldots, n$ правая часть одного из этих уравнений равна 1. Но система неоднородных линейных уравнений имеет решение только тогда, когда ее определитель отличен от нуля, что и доказывает наше утверждение. После этого уравнение (61) дает $X''^0 = K_{x''x'} K_{x'}^{-1} X'^0$, или

$$X'' = m'' + K_{x''x'} K_{x'}^{-1} (X' - m').$$

Отсюда следует, что условная плотность вектора X'' при данном значении x' вектора X' определяется формулой

$$f_2(x'' \mid x') = \delta(x'' - m'' - K_{x''x'} K_{x'}^{-1} (x' - m')).$$

Подставив это выражение и выражение (58) плотности $f_1(x')$ в формулу, найдем плотность случайного вектора X:

$$f(x)=\frac{1}{\sqrt{(2\pi)^r\,|K_{x'}|}}\exp\left\{-\frac{1}{2}(x'-m')^T K_{x'}^{-1}(x'-m')\right\}\times$$
$$\times\,\delta\left(x''-m''-K_{x''x'}K_{x'}^{-1}(x'-m')\right). \blacktriangleleft \quad (62)$$

Распределение вероятностей, определяемое этой формулой, называется *вырожденным нормальным распределением*.

Так как δ-функцию всегда можно рассматривать как нормальную плотность, соответствующую ковариационной матрице, все элементы которой равны 0, то для вырожденного нормального распределения справедлива доказанная выше теорема, что *распределения и условные распределения всех проекций нормально распределенного случайного вектора на любые подпространства, образованные осями координат, нормальны*.

Однако проекции случайного вектора на любые подпространства, образованные осями координат, могут быть распределены нормально и в том случае, когда распределение этого вектора отлично от нормального.

Пример 23. Плотность n-мерного вектора X определяется формулой

$$f(x)=f_N(x)+g(x),$$

где $f_N(x)$ — любая нормальная плотность, а $g(x)$ — любая функция, удовлетворяющая условиям:

1) $|g(x)|\leqslant f_N(x)$ при всех x;

2) $\int\limits_{-\infty}^{\infty} g(x)\,dx_k=0$ при всех x_i, $i\neq k$ $(k=1,\ \ldots,\ n)$.

Интегрированием по любой координате вектора x убеждаемся в том, что проекции вектора X на все подпространства, образованные осями координат, распределены нормально. Однако распределение вектора X при $g(x)\not\equiv 0$ отлично от нормального.

§ 5. Характеристические функции случайных величин

5.1. Характеристическая функция и ее свойства. *Характеристической функцией* действительной случайной величины X называется математическое ожидание случайной величины $e^{i\lambda^T X}$, рассматриваемое как функция действительной переменной λ.

На основании (3.3) характеристическая функция величины X выражается через ее плотность $f(x)$ формулой

$$g(\lambda)=Me^{i\lambda^T X}=\int_{-\infty}^{\infty} e^{i\lambda^T x} f(x)\,dx. \qquad (63)$$

Эта формула определяет характеристическую функцию как для скалярной, так и для векторной величины X. Размерность вектора λ, очевидно, всегда совпадает с размерностью случайного вектора X.

Так как $|e^{i\lambda^T X}|=1$ при всех действительных λ, то любая действительная случайная величина имеет характеристическую функцию.

Изучим свойства характеристических функций.

1) Из (63) следует, что *характеристическая функция непрерывна и что* $|g(\lambda)|\leqslant 1$, $g(0)=1$, $g(-\lambda)=\overline{g(\lambda)}$.

2) *Характеристическая функция положительно определенна*: для любых значений $\lambda_1, \ldots, \lambda_N$ переменной λ и любых комплексных $\xi_1, \ldots, \xi_N$

$$\sum_{p,\,q=1}^{N} g(\lambda_p-\lambda_q)\,\xi_p\bar{\xi}_q \geqslant 0. \qquad (64)$$

Из определения характеристической функции и свойств математических ожиданий следует

$$\sum_{p,\,q=1}^{N} g(\lambda_p-\lambda_q)\,\xi_p\bar{\xi}_q=\sum_{p,\,q=1}^{N} Me^{i(\lambda_p^T-\lambda_q^T)X}\xi_p\bar{\xi}_q= \\ =M\left|\sum_{p=1}^{N} e^{i\lambda_p^T X}\xi_p\right|^2 \geqslant 0.$$

Можно доказать, что любая непрерывная положительно определенная функция $g(\lambda)$, $g(0)=1$, может быть характеристической функцией случайной величины [51, 119].

3) *Характеристическая функция* $g_2(\mu)$ *величины* $Y=AX+a$, *полученной в результате линейного преобразования величины* X, *выражается через характеристическую функцию* $g_1(\lambda)$ *величины* X *формулой*

$$g_2(\mu)=e^{i\mu^T a} g_1(A^T\mu). \qquad (65)$$

По определению

$$g_2(\mu)=Me^{i\mu^T Y}=Me^{i\mu^T(AX+a)}=e^{i\mu^T a}Me^{i\mu^T AX}= \\ =e^{i\mu^T a}Me^{i(A^T\mu)^T X}=e^{i\mu^T a}g_1(A^T\mu).$$

4) Из (65), в частности, следует, что *характеристическая функция проекции случайного вектора на любое подпространство G равна сужению его характеристической функции на это подпространство.*

В самом деле, если $a=0$ и A — матрица проектирования на G, то $A^T=A$, $A\lambda=\lambda$ для любого $\lambda\in G$ и $A\lambda=0$ для любого вектора λ, ортогонального к G.

В частном случае, когда подпространство G образовано некоторыми единичными координатными векторами, для нахождения характеристической функции проекции вектора X на подпространство G следует в выражении $g_1(\lambda)$ положить равными нулю все координаты вектора λ в дополнительном подпространстве.

5) *Характеристическая функция $g(\lambda)$ суммы независимых случайных величин $X_1, \ldots, X_n$ равна произведению их характеристических функций $g_k(\lambda)$ $(k=1, \ldots, n)$:*

$$g(\lambda)=\prod_{k=1}^{n} g_k(\lambda). \tag{66}$$

Действительно, по теореме умножения математических ожиданий (29)

$$g(\lambda)=Me^{i\eta^T(X_1+\ldots+X_n)}=\prod_{k=1}^{n} Me^{i\lambda^T X_k}.$$

6) *Если $X_1, \ldots, X_n$ — независимые случайные величины, то характеристическая функция $g(\lambda)$, $\lambda=[\lambda_1\ldots\lambda_n]^T$, составного случайного вектора $X=[X_1\ldots X_n]^T$ равна произведению характеристических функций $g_k(\lambda_k)$ $(k=1, \ldots, n)$ величин $X_1, \ldots, X_n$:*

$$g(\lambda)=\prod_{k=1}^{n} g_k(\lambda_k). \tag{67}$$

Действительно, по теореме умножения математических ожиданий (29)

$$g(\lambda)=Me^{i\lambda^T X}=Me^{i(\lambda_1^T X_1+\ldots+\lambda_n^T X_n)}=\prod_{k=1}^{n} Me^{i\lambda_k^T X_k}.$$

Наоборот, если совместная характеристическая функция величин $X_1, \ldots, X_n$ (составного вектора $X=[X_1, \ldots, X_n]^T$) выражается формулой (67), то величины $X_1, \ldots, X_n$ независимы. Это следует из формулы (72), выражающей плотность через характеристическую функцию.

Таким образом, *случайные величины $X_1, \ldots, X_n$ независимы тогда и только тогда, когда их совместная характеристическая функция равна произведению их характеристических функций.*

Пример 24. Характеристическая функция дискретной величины X с возможными значениями $x_1, \ldots, x_N$ и их вероятностями $p_1, \ldots, p_N$ определяется формулой

$$g(\lambda) = \sum_{\nu=1}^{N} p_\nu e^{i\lambda^T x_\nu}.$$

Пример 25. Для биномиального распределения

$$g(\lambda) = \sum_{m=0}^{n} C_n^m p^m q^{n-m} e^{i\lambda m} = (pe^{i\lambda} + q)^n.$$

Пример 26. Для распределения Пуассона

$$g(\lambda) = \sum_{m=0}^{\infty} \frac{\mu^m}{m!} e^{-\mu} e^{i\lambda m} = e^{-\mu} \sum_{m=0}^{\infty} \frac{(\mu e^{-\lambda})^m}{m!} = e^{\mu (e^{i\lambda} - 1)}.$$

Пример 27. Для равномерного распределения в интервале (a, b)

$$g(\lambda) = \int_a^b \frac{e^{i\lambda x} dx}{b-a} = \frac{e^{i\lambda b} - e^{i\lambda a}}{i\lambda (b-a)}.$$

Пример 28. Характеристическая функция скалярной нормально распределенной случайной величины X определяется формулой

$$g(\lambda) = \frac{1}{\sqrt{2\pi D}} \int_{-\infty}^{\infty} e^{i\lambda x - \frac{(x-m)^2}{2D}} dx = \frac{e^{i\lambda m}}{\sqrt{2\pi D}} \int_{-\infty}^{\infty} e^{i\lambda t - \frac{t^2}{2D}} dt.$$

Применив формулу (1) приложения 2, получаем

$$g(\lambda) = e^{i\lambda m - \frac{1}{2} D\lambda^2}. \qquad (68)$$

Пример 29. Для γ-распределения

$$g(\lambda) = \int_0^\infty e^{i\lambda x} \frac{k^{\mu+1}}{\Gamma(\mu+1)} x^\mu e^{-kx} dx =$$

$$= \frac{k^{\mu+1}}{(k-i\lambda)^{\mu+1} \Gamma(\mu+1)} \int_0^{(k-i\lambda)\infty} y^\mu e^{-y} dy,$$

где интегрирование производится по лучу l в плоскости комплексной переменной y, выходящему из начала координат в направлении вектора $k - i\lambda$. Применив теорему Коши о равенстве нулю интеграла по замкнутому контуру от аналитической функции, не имеющей особых точек в области, ограниченной контуром, убеждаемся в том, что интегрирование по лучу l можно заменить интегрированием по

положительной части действительной оси. Тогда в силу определения (2.16) гамма-функции интеграл будет равен $\Gamma(\mu+1)$ и мы получим

$$g(\lambda)=k^{\mu+1}(k-i\lambda)^{-\mu-1}. \tag{69}$$

Пример 30. В частном случае показательного распределения $\mu=0$ и формула (69) принимает вид

$$g(\lambda)=\frac{k}{k-i\lambda}.$$

Пример 31. В другом частном случае χ^2-распределения $\mu=n/2-1$, $k=1/2$ и формула (69) дает

$$g(\lambda)=(1-2i\lambda)^{-n/2}.$$

Пример 32. Для n-мерного нормально распределенного случайного вектора

$$g(\lambda)=\frac{1}{\sqrt{(2\pi)^n|K|}}\int_{-\infty}^{\infty}\exp\left\{i\lambda^T x-\frac{1}{2}(x^T-m^T)K^{-1}(x-m)\right\}dx.$$

Применив для вычисления интеграла формулу (2) приложения 2, получим

$$g(\lambda)=\exp\left\{i\lambda^T m-\frac{1}{2}\lambda^T K\lambda\right\}. \tag{70}$$

5.2. Выражение плотности через характеристическую функцию. *Характеристическая функция случайной величины полностью и однозначно определяет ее распределение.* Мы докажем это утверждение для случайных величин, имеющих плотность, возможно содержащую линейную комбинацию δ-функций. Доказательство в общем случае читатель найдет, например, в [51, 119].

▶ Плотность непрерывной случайной величины неотрицательна и интегрируема. Поэтому, предположив, что она непрерывна всюду, кроме, может быть, конечного множества точек (линий, поверхностей, гиперповерхностей в случае векторной величины), можем представить ее интегралом Фурье:

$$f(x)=\frac{1}{(2\pi)^n}\lim_{\Lambda\to\infty}\int_{-\Lambda}^{\Lambda}e^{-i\lambda^T x}g(\lambda)\,d\lambda. \tag{71}$$

Эта формула определяет $f(x)$ во всех точках ее непрерывности, т. е. при всех x, кроме, может быть, некоторого множества точек нулевой суммарной меры. Следовательно, она полностью определяет распределение непрерывной случайной величины.

Формулу (71) обычно записывают в виде

$$f(x)=\frac{1}{(2\pi)^n}\int\limits_{-\infty}^{\infty} e^{-i\lambda^T x} g(\lambda)\, d\lambda, \qquad (72)$$

где интеграл понимается в смысле главного значения Коши в случае, когда он сходится не абсолютно [69, 109].

Формула (72) определяет плотность и в случае дискретной или непрерывно-дискретной величины X, так как в силу формулы (15) приложения 1 δ-функция представима интегралом Фурье:

$$\delta(x)=\frac{1}{(2\pi)^n}\int\limits_{-\infty}^{\infty} d\lambda \int\limits_{-\infty}^{\infty} e^{i\lambda^T (u-x)}\delta(u)\, du.$$

Таким образом, формула (72) определяет плотность, а следовательно, и распределение по данной характеристической функции для всех трех типов случайных величин, встречающихся в задачах практики. ◀

5.3. Связь между характеристической функцией и моментами. Чтобы установить связь между характеристической функцией и моментами случайной величины, рассмотрим сначала случай скалярной величины. Ее характеристическая функция в соответствии с (63) определяется формулой

$$g(\lambda)=\int\limits_{-\infty}^{\infty} e^{i\lambda x} f(x)\, dx.$$

▶ Предположим, что существует момент r-го порядка α_r величины X. В этом случае существует и производная

$$g^{(r)}(\lambda)=i^r\int\limits_{-\infty}^{\infty} x^r e^{i\lambda x} f(x)\, dx=i^r M X^r e^{i\lambda X}.$$

Положив здесь $\lambda=0$, находим выражение для момента α_r через характеристическую функцию:

$$\alpha_r=i^{-r}g^{(r)}(0). \qquad (73)$$

Если величина X имеет моменты до порядка ν включительно, то на основании (73) характеристическую функцию $g(\lambda)$ можно представить формулой Маклорена:

$$g(\lambda)=1+\sum_{r=1}^{\nu}\frac{i^r\alpha_r}{r!}\lambda^r+R_\nu, \qquad (74)$$

где R_ν — остаточный член.

Заметив, что $e^{-i\lambda m_x}g(\lambda)$ представляет собой характеристическую функцию центрированной случайной величины $X^0 = X - m_x$, из (73) получим

$$\mu_r = i^{-r}\left[\frac{d^r}{d\lambda^r} e^{-i\lambda m_x} g(\lambda)\right]_{\lambda=0}. \tag{75}$$

Если случайная величина X имеет конечные центральные моменты до порядка ν включительно, то функцию $e^{-i\lambda m_x}g(\lambda)$ можно представить формулой Маклорена:

$$e^{-i\lambda m_x} g(\lambda) = 1 + \sum_{r=2}^{\nu} \frac{i^r \mu_r}{r!}\lambda^r + R'_\nu. \tag{76}$$

Совершенно так же для n-мерного случайного вектора X получаем формулы для моментов $\alpha_{h_1, \ldots, h_n}$ и $\mu_{h_1, \ldots, h_n}$, $h_1 + \ldots + h_n = r$:

$$\alpha_{h_1, \ldots, h_n} = i^{-r}\left[\frac{\partial^r g(\lambda)}{\partial\lambda_1^{h_1}\ldots\partial\lambda_n^{h_n}}\right]_{\lambda=0}, \tag{77}$$

$$\mu_{h_1, \ldots, h_n} = i^{-r}\left[\frac{\partial^r}{\partial\lambda_1^{h_1}\ldots\partial\lambda_n^{h_n}} e^{-i\lambda^T m_x} g(\lambda)\right]_{\lambda=0}, \tag{78}$$

и выражения функций $g(\lambda)$ и $e^{-i\lambda^T m_x}g(\lambda)$ через моменты:

$$g(\lambda) = 1 + \sum_{r=1}^{\nu} i^r \sum_{h_1+\ldots+h_n=r} \frac{\alpha_{h_1, \ldots, h_n}}{h_1!\ldots h_n!}\lambda_1^{h_1}\ldots\lambda_n^{h_n} + R_\nu, \tag{79}$$

$$e^{-i\lambda^T m_x} g(\lambda) = \\ = 1 + \sum_{r=2}^{\nu} i^r \sum_{h_1+\ldots+h_n=r} \frac{\mu_{h_1, \ldots, h_n}}{h_1!\ldots h_n!}\lambda_1^{h_1}\ldots\lambda_n^{h_n} + R'_\nu, \tag{80}$$

где равенство под знаком внутренней суммы указывает, что суммирование распространяется на все значения $h_1, \ldots, h_n = 0, 1, \ldots, r$, сумма которых равна r. ◀

Полученные соотношения дают два простых и удобных способа вычисления моментов случайной величины: способ дифференцирования характеристической функции и способ разложения ее по степеням λ.

Пример 33. Для скалярной нормально распределенной случайной величины X на основании (68) имеем

$$e^{-i\lambda m_x} g(\lambda) = e^{-\frac{1}{2}D\lambda^2} = \sum_{s=0}^{\infty} \frac{(-1)^s D_x^s}{2^s s!}\lambda^{2s} = \sum_{s=0}^{\infty} \frac{i^{2s} D_x^s}{2^s s!}\lambda^{2s}.$$

Сравнив эту формулу с (76), находим

$$\mu_{2s-1}=0, \qquad \mu_{2s}=\frac{(2s)!}{2^s s!} D_x^s.$$

Этот результат совпадает с полученным в § 3.6. В данном случае способ разложения характеристической функции в ряд значительно проще, чем непосредственное вычисление моментов.

Пример 34. Для нормально распределенного n-мерного случайного вектора на основании (70) имеем

$$e^{-i\lambda^T m_x} g(\lambda)=e^{-\frac{1}{2}\lambda^T K_x \lambda}=\sum_{s=0}^{\infty}\frac{(-1)^s}{2^s s!}(\lambda^T K_x \lambda)^s=$$

$$=\sum_{s=0}^{\infty}\frac{i^{2s}}{2^s s!}\sum_{p_1, q_1, \ldots, p_s, q_s=1}^{n} k_{p_1 q_1}\ldots k_{p_s q_s}\lambda_{p_1}\lambda_{q_1}\ldots\lambda_{p_s}\lambda_{q_s}. \quad (81)$$

Отсюда видно, что все центральные моменты нечетного порядка нормально распределенного случайного вектора равны нулю. Для центрального момента четного порядка $h_1+\ldots+h_n=2s$ из сравнения (80) и (81) вытекает формула

$$\mu_{h_1, \ldots, h_n}=\frac{h_1!\ldots h_n!}{2^s s!}\sum k_{p_1 q_1}\ldots k_{p_s q_s}, \quad (82)$$

где сумма распространена на все возможные различные перестановки $2s$ индексов $p_1, q_1, \ldots, p_s, q_s$, из которых h_1 индексов равны 1, h_2 равны 2, ..., h_n равны n. Очевидно, что число таких перестановок, а следовательно и число слагаемых в сумме (82), равно $(2s)!/(h_1!\ldots h_n!)$. Положив в (82) последовательно $s=2$, $h_1=4$, $h_k=0$ при $k>1$; $h_1=3$; $h_2=1$, $h_k=0$ при $k>2$; $h_1=h_2=2$, $h_k=0$ при $k>2$; $h_1=2$, $h_2=h_3=1$, $h_k=0$ при $k>3$; $h_1=h_2=h_3=h_4=1$, $h_k=0$ при $k>4$, получим следующие формулы для центральных моментов четвертого порядка:

$$\begin{aligned} &\mu_{40\ldots0}=3k_{11}^2, \qquad \mu_{310\ldots0}=3k_{11}k_{12},\\ &\mu_{220\ldots0}=k_{11}k_{22}+2k_{12}^2,\\ &\mu_{2110\ldots0}=k_{11}k_{23}+2k_{12}k_{13},\\ &\mu_{11110\ldots0}=k_{12}k_{34}+k_{13}k_{24}+k_{14}k_{23}. \end{aligned} \quad (83)$$

Формулы (82) и (83) выражают центральные моменты нормально распределенного случайного вектора через элементы его ковариационной матрицы.

Заметим, что приведенный вывод формулы (82) для моментов нормально распределенного случайного вектора при помощи разложения его характеристической функции в ряд крайне прост, в то время как непосредственное вычисление моментов по формулам § 3.5 было бы весьма громоздким.

Если распределение случайной величины полностью сосредоточено в ограниченной области, то ее моменты всех порядков существуют и полностью определяют ее распределение.

▶ Для доказательства достаточно заметить, что если $|X_k| < l_k$ $(k = 1, \ldots, n)$, то $|\alpha_{h_1, \ldots, h_n}| < l_1^{h_1} \ldots l_n^{h_n}$, вследствие чего члены ряда Маклорена характеристической функции по модулю меньше соответствующих членов сходящегося при всех $\lambda_1, \ldots, \lambda_n$ ряда

$$1 + \sum_{r=1}^{\infty} \sum_{h_1 + \ldots + h_n = r} \frac{l_1^{h_1} \ldots l_n^{h_n}}{h_1! \ldots h_n!} \lambda_1^{h_1} \ldots \lambda_n^{h_n} = e^{-\lambda^T l},$$

где $l = [l_1 \ldots l_n]^T$. ◀

5.4. Семиинварианты. В некоторых случаях удобно пользоваться разложением по степеням λ не самой характеристической функции, а ее логарифма.

Если существуют моменты скалярной случайной величины X до порядка ν включительно, то на основании формулы Маклорена

$$\ln g(\lambda) = \sum_{r=1}^{\nu} \frac{i^r \varkappa_r}{r!} \lambda^r + R''_\nu. \tag{84}$$

Коэффициенты

$$\varkappa_r = i^{-r} \left[\frac{d^r \ln g(\lambda)}{d\lambda^r} \right]_{\lambda=0} \qquad (r = 1, \ldots, \nu) \tag{85}$$

называются *семиинвариантами* или *кумулянтами* случайной величины X. Выразив производные $\ln g(\lambda)$ через соответствующие производные характеристической функции, можно на основании (73) выразить семиинварианты случайной величины через ее моменты, и наоборот. Мы предоставляем читателю самостоятельно найти эти зависимости. Если воспользоваться формулой

$$\ln g(\lambda) = i m_x \lambda + \ln \{e^{-i\lambda m_x} g(\lambda)\} \tag{86}$$

и принять во внимание (75), то получим выражение семиинвариантов случайной величины через ее математическое ожидание и центральные моменты:

$$\begin{aligned} &\varkappa_1 = m_x, \quad \varkappa_2 = \mu_2 = D_x, \quad \varkappa_3 = \mu_3, \\ &\varkappa_4 = \mu_4 - 3\mu_2^2, \quad \varkappa_5 = \mu_5 - 10\mu_2\mu_3, \\ &\varkappa_6 = \mu_6 - 15\mu_2\mu_4 - 10\mu_3^2 + 30\mu_2^3, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{87}$$

Если существуют моменты n-мерного случайного вектора X до порядка ν включительно, то формула

Маклорена дает

$$\ln g(\lambda)=\sum_{r=1}^{\nu} i^r \sum_{h_1+\ldots+h_n=r} \frac{\varkappa_{h_1,\ldots,h_n}}{h_1!\ldots h_n!} \lambda_1^{h_1}\ldots\lambda_n^{h_n}+R''_\nu. \quad (88)$$

Коэффициенты

$$\varkappa_{h_1,\ldots,h_n}=i^{-r}\left[\frac{\partial^r \ln g(\lambda)}{\partial\lambda_1^{h_1}\ldots\partial\lambda_n^{h_n}}\right]_{\lambda=0} \qquad (h_1+\ldots+h_n=r) \quad (89)$$

называются *семиинвариантами* или *кумулянтами* случайного вектора X. Ясно, что семиинварианты порядка r выражаются через моменты до порядка r включительно, и наоборот.

Формулы (68) и (70) показывают, что *для нормально распределенной случайной величины все семиинварианты выше второго порядка равны нулю*. Следовательно, семиинварианты случайной величины, начиная с третьего порядка, характеризуют отклонение ее распределения от нормального.

Семиинварианты обладают одним интересным свойством: *семиинварианты суммы независимых случайных величин равны суммам соответствующих семиинвариантов слагаемых*. Это следует из свойства 5) характеристических функций.

5.5. Порядок остаточных членов в разложениях. Остаточные члены в (79), (80) и (88) имеют порядок $o(|\lambda|^\nu)$. Мы докажем это, предполагая для простоты, что величина X имеет конечные моменты $(\nu+1)$-го порядка (хотя достаточно существования конечных моментов ν-го порядка).

► Представив $e^{i\lambda^T X}$ формулой Маклорена и имея в виду, что $|\lambda^T X|\leqslant|\lambda||X|$, получаем

$$g(\lambda)=1+i\lambda^T m_x+\sum_{r=2}^{\nu}\frac{i^r}{r!}M(\lambda^T X)^r+R_\nu, \quad (90)$$

где

$$|R_\nu|=\frac{1}{(\nu+1)!}\left|M(\lambda^T X)^{\nu+1}e^{i\theta\lambda^T X}\right|\leqslant$$

$$\leqslant\frac{|\lambda|^{\nu+1}}{(\nu+1)!}M|X|^{\nu+1}=o(|\lambda|^\nu).$$

Так как $e^{-i\lambda^T m_x} g(\lambda)$ есть характеристическая функция центрированной случайной величины X^0, то из доказанного следует, что и $R'_\nu = o(|\lambda|^\nu)$.

Наконец, подставив в формулу

$$\ln(1+z) = \sum_{r=1}^{\nu} (-1)^{r+1} \frac{z^r}{r} + o(z^\nu) \tag{91}$$

выражение $z = g(\lambda) - 1$ из (90), убеждаемся в том, что и $R''_\nu = o(|\lambda|^\nu)$. ◀

В частности, положив в (90) и (91) $\nu = 2$, получаем для характеристической функции случайной величины X с конечными моментами третьего порядка

$$\ln g(\lambda) = i\lambda^T m_x - \frac{1}{2} \lambda^T K_x \lambda + o(|\lambda|^2) \tag{92}$$

(эта формула справедлива и в случае, когда X имеет только моменты первого и второго порядков).

Глава 5

ФУНКЦИИ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 1. Моменты функций случайных величин

1.1. Точные формулы для первых и вторых моментов. В п. 3.1.3 и п. 3.3.5 мы научились находить первые и вторые моменты линейных функций случайных величин. Для определения этих моментов достаточно знать первые и вторые моменты величин-аргументов и не требуется знать распределения величин-аргументов. Иначе обстоит дело с нелинейными функциями случайных величин.

Предположим, что случайный вектор Y (в общем случае комплексный) представляет собой данную функцию действительного случайного вектора X, имеющего плотность $f(x)$, $Y=\varphi(X)$. Моменты величины Y можно определить по формуле (3.3) для математического ожидания функции случайной величины. Пользуясь этой формулой и представляя векторы в виде матриц-столбцов, на основании определений (3.35) получаем

$$m_y = MY = M\varphi(X) = \int_{-\infty}^{\infty} \varphi(x) f(x)\,dx, \qquad (1)$$

$$\Gamma_y = MYY^* = M\varphi(X)\varphi(X)^* = \int_{-\infty}^{\infty} \varphi(x)\varphi(x)^* f(x)\,dx, \qquad (2)$$

$$K_y = MY^0Y^{0*} = M[\varphi(X)-m_y][\varphi(X)^*-m_y^*] = \\ = \int_{-\infty}^{\infty} [\varphi(x)-m_y][\varphi(x)^*-m_y^*] f(x)\,dx. \qquad (3)$$

Точно так же по формуле (3.3) определяется взаимный второй момент и взаимная ковариационная матрица двух случайных векторов, представляющих собой данные функции $Y=\varphi(X)$, $Z=\psi(X)$ действительного случайного

вектора X:

$$\Gamma_{yz} = MYZ^* = M\varphi(X)\psi(X)^* = \int_{-\infty}^{\infty} \varphi(x)\psi(x)^* f(x)\,dx, \quad (4)$$

$$K_{yz} = MY^0Z^{0*} = \int_{-\infty}^{\infty} [\varphi(x) - m_y][\psi(x)^* - m_z^*] f(x)\,dx. \quad (5)$$

Формулы (1)—(5) применимы и к функциям комплексного случайного вектора. В этом случае под X следует понимать $2n$-мерный действительный случайный вектор, координатами которого служат действительные и мнимые части координат n-мерного комплексного случайного вектора-аргумента.

В частном случае одномерных векторов Y и Z формулы (1)—(5) определяют математические ожидания и моменты второго порядка скалярных функций случайных величин.

Формулы (1)—(5) показывают, что для определения математических ожиданий и моментов второго порядка нелинейных функций случайных величин необходимо в общем случае знать плотность величин-аргументов.

Предлагаем читателю самостоятельно доказать, что в частном случае линейных функций $\varphi(X)$ и $\psi(X)$ из формул (1), (3) и (5) вытекают соответственно формулы (3.49), (3.50) и (3.55) для математических ожиданий и ковариационных матриц линейных функций случайного вектора X.

Пример 1. Шарики для шариковых клапанов изготовляются с ошибками в радиусе, распределенными по нормальному закону с математическим ожиданием r_0 и средним квадратическим отклонением σ (конечно, $\sigma \ll r_0$). Найти математическое ожидание и дисперсию площади диаметрального сечения шарика.

Площадь S диаметрального сечения шарика выражается через его случайный радиус известной формулой $S = \pi R^2$. Следовательно, случайная величина S в данном случае является нелинейной функцией случайной величины R. Поэтому для определения m_s и D_s следует воспользоваться формулами (1) и (3). Применяя формулу (1), будем иметь

$$m_s = \int_{-\infty}^{\infty} \pi r^2 f(r)\,dr = \pi \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} r^2 e^{-\frac{1}{2}\left(\frac{r-r_0}{\sigma}\right)^2} dr.$$

Последний интеграл вместе с множителем $1/\sigma\sqrt{2\pi}$ представляет собой начальный момент второго порядка случайной величины R, который на основании формулы (3.29) равен $r_0^2 + \sigma^2$. Следовательно,

$$m_s = \pi(r_0^2 + \sigma^2).$$

После этого формула (3) дает

$$D_s = \pi^2 \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (r^2 - r_0^2 - \sigma^2)^2 e^{-\left(\frac{r-r_0}{\sigma}\right)^2} dr.$$

Вычислив интеграл (что можно сделать, раскрыв скобки под знаком интеграла и путем интегрирования по частям избавившись от четвертой степени r), получим

$$D_s = 2\pi^2\sigma^2 (2r_0^2 + \sigma^2).$$

Пример 2. Найти математическое ожидание и дисперсию случайной величины

$$Y = X^3 - kX,$$

где X — случайная величина, равномерно распределенная на интервале $(-a, a)$.

Плотность величины X в интервале $(-a, a)$ равна $1/(2a)$, а вне этого интервала равна нулю. Поэтому, применяя формулу (1), получаем

$$m_y = \frac{1}{2a} \int_{-a}^{a} (x^3 - kx)\, dx = 0.$$

После этого формула (3) дает

$$D_y = \frac{1}{2a} \int_{-a}^{a} (x^3 - kx)^2\, dx = \frac{15a^4 - 42ka^2 + 35k^2}{105} a^2.$$

Пример 3. В условиях примера 2 найти ковариацию случайных величин

$$Y = X^3 - kX, \qquad Z = X^3 + lX.$$

Сначала находим по формуле (1) математическое ожидание случайной величины Z, которое оказывается равным нулю. После этого по формуле (5) находим

$$k_{yz} = \frac{1}{2a} \int_{-a}^{a} (x^3 - kx)(x^3 + lx)\, dx = \frac{15a^4 - 21ka^2 + 21la^2 - 35kl}{105} a^2.$$

1.2. Метод линеаризации. Сложность вычисления математических ожиданий, дисперсий и ковариаций нелинейных функций случайных величин приводит к естественному желанию пожертвовать точностью ради простоты и найти простые приближенные формулы, подобные формулам для моментов линейных функций случайных величин. Для получения таких приближенных формул обычно применяют *метод линеаризации*, который состоит в замене нелинейных функций достаточно близкими к ним линейными функциями.

Рассмотрим сначала случай скалярной функции скалярной случайной величины X. В этом случае линеаризация нелинейной функции $\varphi(X)$ сводится к замене кривой $y=\varphi(x)$ некоторой прямой $y=ax+b$. Если удастся подобрать прямую, достаточно близкую к кривой в области практически возможных значений случайной величины X (в случае нормального распределения случайной величины X в интервале $(m_x-3\sigma_x,\ m_x+3\sigma_x)$), то можно рассчитывать на то, что математическое ожидание и дисперсия соответствующей линейной функции случайной величины X будут близкими к математическому ожиданию и дисперсии нелинейной функции.

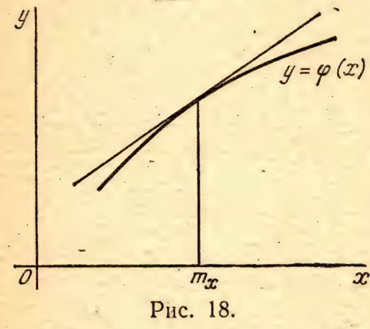


Рис. 18.

Посмотрим теперь, как следует выбрать прямую, заменяющую данную кривую. Так как плотность случайной величины обычно имеет бо́льшие значения вблизи математического ожидания и меньшие вдали от математического ожидания (возможные значения случайной величины расположены «гуще» около математического ожидания и «реже» вдали от него), то естественно добиваться более высокой точности приближения прямой к заданной кривой около математического ожидания случайной величины-аргумента, т. е. около точки $x=m_x$. Поэтому наиболее естественным способом линеаризации для наших целей является замена кривой $y=\varphi(x)$ отрезком касательной к ней в точке $x=m_x$ (рис. 18). При такой замене истинная нелинейная зависимость случайной величины Y от X заменяется приближенной линейной зависимостью

$$Y \approx \varphi(m_x) + \varphi'(m_x) X^0. \tag{6}$$

Само собой разумеется, что для возможности такой замены необходимо существование производной функции $\varphi(x)$ в точке $x=m_x$. С аналитической точки зрения замена нелинейной функции $\varphi(x)$ линейной функцией (6) равноценна замене приращения функции $\varphi(x)$ в точке m_x ее дифференциалом.

Аналогично получаем формулу (6) в случае векторной функции случайного вектора X, дифференцируемой в точке m_x. В этом случае $\varphi'(m_x)$ следует понимать как

матрицу частных производных всех координат вектора $\varphi(x)$ по всем координатам вектора x в точке $x=m_x$:

$$\varphi'(m_x)=\begin{bmatrix} \frac{\partial\varphi_1}{\partial x_1} & \frac{\partial\varphi_1}{\partial x_2} & \cdots & \frac{\partial\varphi_1}{\partial x_n} \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial\varphi_s}{\partial x_1} & \frac{\partial\varphi_s}{\partial x_2} & \cdots & \frac{\partial\varphi_s}{\partial x_n} \end{bmatrix}_{x=m_x}. \qquad (7)$$

Заменив функцию $\varphi(X)$ линейной функцией (6), можно применить для вычисления математического ожидания и ковариационной матрицы случайного вектора Y формулы (3.49) и (3.50). В результате получим приближенные формулы

$$m_y\approx\varphi(m_x),\quad K_y\approx\varphi'(m_x)\,K_x\varphi'(m_x)^*. \qquad (8)$$

Формулы (8), справедливые как для действительных, так и для комплексных скалярных или векторных случайных величин X и Y, позволяют достаточно просто, без сложных вычислений приближенно определять математические ожидания и моменты второго порядка любых нелинейных функций случайных величин, непрерывных вместе со своими первыми производными. Эти формулы тем точнее, чем меньше отклоняется функция $\varphi(x)$ от заменяющей ее линейной функции (7) в области практически возможных значений величины-аргумента X. Чем сильнее отклоняется функция $\varphi(x)$ от линейной функции (7), тем меньше должны быть дисперсии координат вектора X для того, чтобы можно было пользоваться формулами (8). Если $\varphi(x)$ слабо отклоняется от линейной функции (т. е. $\varphi'(x)$ изменяется медленно с изменением x), то формулы (8) могут быть достаточно точными и при больших дисперсиях координат вектора X. Если $\varphi(x)$ значительно отличается от линейной ($\varphi'(x)$ изменяется быстро), то формулами (8) можно пользоваться только при достаточно малых дисперсиях координат вектора X.

Пример 4. В условиях примера 1 найти математическое ожидание и дисперсию площади диаметрального сечения шарика методом линеаризации.

В данном случае $S=\varphi(R)=\pi R^2$. Поэтому $\varphi'(r)=2\pi r$ и формулы (8) дают

$$m_s\approx\varphi(m_r)=\varphi(r_0)=\pi r_0^2,$$
$$D_s\approx|\varphi'(m_r)|^2\sigma^2=|\varphi'(r_0)|^2\sigma^2=4\pi r_0^2\sigma^2.$$

Сравнив эти формулы с точными формулами примера 1, видим, что метод линеаризации в данном случае дает хорошую точность, если

$\sigma^2 \ll r_0^2$, т. е. если дисперсия радиуса шарика мала по сравнению с квадратом его математического ожидания. Например, при $\sigma \approx 0{,}1 r_0$ относительная ошибка определения m_s методом линеаризации составляет 1%, а относительная ошибка определения дисперсии D_s — 0,5%.

Пример 5. Решить пример 2 методом линеаризации.

В данном случае $\varphi(x) = x^3 - kx$, $\varphi'(x) = 3x^2 - k$, и формулы (8) дают

$$m_y = \varphi(m_x) = \varphi(0) = 0,$$
$$D_y = |\varphi'(m_x)|^2 D_x = |\varphi'(0)|^2 D_x = k^2 D_x.$$

Вспомнив, что дисперсия случайной величины, равномерно распределенной на интервале длины $2a$, равна $a^2/3$ (пример 3.3), получим окончательно $D_y \approx k^2a^2/3$. Сравнив этот результат с точной формулой примера 2, видим, что в данном случае метод линеаризации дает хорошую точность, если $a^2 \ll k$.

Пример 6. Решить пример 3 методом линеаризации.

В данном случае $\varphi(x)$ представляет собой двумерный вектор с координатами $\varphi_1(x) = x^3 - kx$, $\varphi_2(x) = x^3 + lx$,

$$\varphi'(m_x) = \varphi'(0) = \begin{bmatrix} \varphi_1'(0) \\ \varphi_2'(0) \end{bmatrix} = \begin{bmatrix} -k \\ l \end{bmatrix},$$

и вторая формула (8) дает

$$\begin{bmatrix} D_y & k_{yz} \\ k_{yz} & D_z \end{bmatrix} \approx \begin{bmatrix} -k \\ l \end{bmatrix} D_x \begin{bmatrix} -k & l \end{bmatrix} = \begin{bmatrix} k^2D_x & -klD_x \\ -klD_x & l^2D_x \end{bmatrix},$$

откуда находим

$$k_{yz} = -klD_x = -\frac{kla^2}{3}.$$

Сравнив эту формулу с точной формулой примера 3, видим, что в данном случае метод линеаризации дает хорошую точность, если $a^2 \ll k, l$.

Пример 7. Ролики для подшипников изготовляются с ошибками, приводящими к тому, что каждый ролик представляет собой эллиптический цилиндр с полуосями $r_0 + X_1$ и $r_0 + X_2$ высоты $l_0 + X_3$, где r_0 и l_0 — номинальные значения радиуса и высоты ролика, а X_1, X_2, X_3 — случайные ошибки производства, представляющие собой координаты нормально распределенного случайного вектора с нулевым математическим ожиданием. Найти математическое ожидание и дисперсию объема ролика.

В данном случае $V = \varphi(X) = \pi(r_0 + X_1)(r_0 + X_2)(l_0 + X_3)$,

$$\varphi'(m_x) = \left[\frac{\partial\varphi}{\partial x_1} \; \frac{\partial\varphi}{\partial x_2} \; \frac{\partial\varphi}{\partial x_3}\right]_{x=0} = \left[\pi r_0 l_0 \;\; \pi r_0 l_0 \;\; \pi r_0^2\right],$$

и формулы (8) дают

$$m_v \approx \varphi(m_x) = \varphi(0) = \pi r_0^2 l_0,$$

$$D_v \approx \left[\pi r_0 l_0 \pi r_0 l_0 \pi r_0^2\right] \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{12} & k_{22} & k_{23} \\ k_{13} & k_{23} & k_{33} \end{bmatrix} \begin{bmatrix} \pi r_0 l_0 \\ \pi r_0 l_0 \\ \pi r_0^2 \end{bmatrix} =$$

$$= \pi^2 r_0^2 \left[(k_{11} + 2k_{12} + k_{22}) l_0^2 + 2(k_{13} + k_{23}) l_0 r_0 + k_{33} r_0^2\right].$$

В частном случае независимых ошибок X_1, X_2, X_3 $k_{12}=k_{13}=k_{33}=0$ и

$$D_v \approx \pi^2 r_0^2 \left[(D_1+D_2)\, l_0^2 + D_3 r_0^2\right],$$

где $D_1=k_{11}$, $D_2=k_{22}$, $D_3=k_{33}$.

Пример 8. Оценить точность определения математического ожидания и дисперсии случайной величины $Y=\varphi(X)$ для случая скалярных X и Y, предполагая, что функция $\varphi(x)$ имеет ограниченную вторую производную в области возможных значений величины X.

§ 2. Функция распределения функции случайного аргумента

2.1. Общий принцип определения распределений функций. Во многих задачах практики, особенно в математической статистике, необходимо уметь находить распределение функции случайонго аргумента. Мы будем решать эту задачу, предполагая все случайные величины, как скалярные, так и векторные, действительными. Способы нахождения распределения функции $Y=\varphi(X)$ случайной величины X основаны на следующем очевидном положении. Чтобы случайная величина Y попала на множество B, необходимо и достаточно, чтобы величина X попала на множество A_B. значений x, которым соответствуют значения функции $\varphi(x)$, принадлежащие множеству B, $A_B=\{x\colon\ \varphi(x)\in B\}$*). На рис. 19 это проиллюстрировано для скалярных величин X и Y. Поэтому вероятность попадания величины $Y=\varphi(X)$ на множество B равна вероятности попадания величины X на множество $A_B=\{x\colon\ \varphi(x)\in B\}$.

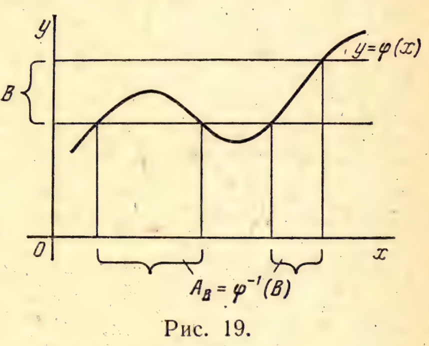


Рис. 19.

2.2. Нахождение функции распределения. В зависимости от того, как выбрать множество B, получаются разные способы нахождения распределения величины Y. В частности, если принять $B=\{Y<y\}$, вероятность попадания величины X на соответствующее множество

*) Такое множество A_B называется *прообразом* множества B (соответствующим функции $\varphi(x)$) и обозначается $\varphi^{-1}(B)$, $A_B=\varphi^{-1}(B)$.

$A_B = A_y = \{x\colon \varphi(x) < y\}$ будет представлять собой функцию распределения случайной величины $Y = \varphi(X)$.

Очевидно, что для возможности определения функции распределения величины $Y = \varphi(X)$ необходимо и достаточно, чтобы функция $\varphi(x)$ удовлетворяла единственному условию: при любом y должна быть определена вероятность попадания случайной величины X на множество $A_y = \{x\colon \varphi(x) < y\}$. Такие функции называются *измеримыми*. Функции, с которыми приходится встречаться в задачах практики, всегда бывают не только измеримыми, но по большей части и непрерывными, а часто и дифференцируемыми.

Предположим, что известна плотность $f_1(x)$ случайной величины X. Тогда на основании сформулированного принципа функция распределения $F_2(y)$ случайной величины $Y = \varphi(X)$ определяется формулой

$$F_2(y) = \int_{A_y} f_1(x)\,dx = \int_{\varphi(x) < y} f_1(x)\,dx. \qquad (9)$$

Эта формула справедлива как для скалярных, так и для векторных случайных величин X и Y (включая случаи, когда одна из них скалярная — одномерный вектор, а другая векторная).

Пример 9. Дана плотность $f_1(x, y)$ двумерного случайного вектора (X, Y). Найти распределение отношения координат этого вектора $Z = Y/X$.

Учитывая, что $y/x < z$ при $y < zx$, если $x > 0$, и при $y > zx$, если $x < 0$, находим по формуле (9) функцию распределения величины Z:

$$F_2(z) = \iint_{y/x < z} f_1(x, y)\,dx\,dy = \int_{-\infty}^{0} dx \int_{zx}^{\infty} f_1(x, y)\,dy + \int_{0}^{\infty} dx \int_{-\infty}^{zx} f_1(x, y)\,dy.$$

Дифференцируя эту формулу по z, находим плотность величины Z:

$$f_2(z) = \int_{-\infty}^{\infty} |x|\, f_1(x, zx)\,dx. \qquad (10)$$

В частности, в случае круговой симметрии распределения аргументов $f_1(x, y) = p(x^2 + y^2)$ и формула (10) дает

$$f_2(z) = 2\int_{0}^{\infty} xp\,(x^2(1 + z^2))\,dx.$$

Замена переменных $u = x\sqrt{1+z^2}$ приводит эту формулу к виду $f_2(z) = \frac{2}{1+z^2}\int\limits_0^\infty u p(u^2)\,du$. Интеграл здесь представляет собой постоянную, которая легко определяется из условия равенства единице интеграла от плотности $f_2(z)$ по всей области возможных значений отношения $Z = Y/X$, т. е. от $-\infty$ до ∞. В итоге получаем распределение Коши (пример 3.6): $f_2(z) = \frac{1}{\pi(1+z^2)}$.

Пример 10. Распределение n-мерного случайного вектора X задано плотностью $f_1(x) = p(x^TCx)$, где C — положительно определенная симметричная матрица. Найти распределение квадратичной формы $Y = X^TCX$.

По формуле (9) находим $F_2(y) = \int\limits_{x^TCx<y} p(x^TCx)\,dx$. Любую симметричную положительно определенную матрицу можно ортогональным преобразованием привести к диагональной форме. Обозначим буквой Λ диагональную матрицу, элементами которой служат собственные значения $\lambda_1, \ldots, \lambda_n$ матрицы C. Тогда, имея в виду, что определитель ортогонального преобразования равен единице, получим после соответствующей замены переменных

$$F_2(y) = \int\limits_{u^T\Lambda u<y} p(u^T\Lambda u)\,du.$$

Положим теперь $\sqrt{\Lambda}\,u = \rho\alpha$ *), где ρ — скалярная неотрицательная переменная (модуль радиуса-вектора точки в n-мерном пространстве), а α — n-мерный вектор с единичным модулем, $\alpha^T\alpha = 1$ (координаты $\alpha_1, \ldots, \alpha_n$ вектора α представляют собой направляющие косинусы радиуса-вектора точки n-мерного пространства). Перейдем от переменных интегрирования $u_1, \ldots, u_n$ к переменным $\rho, \alpha_1, \ldots, \alpha_{n-1}$. Якобиан преобразования в силу соотношений $u_p = \rho\alpha_p/\sqrt{\lambda_p}$ и $\alpha^T\alpha = \alpha_1^2 + \ldots + \alpha_n^2 = 1$ равен

$$J = \begin{vmatrix} \frac{\alpha_1}{\sqrt{\lambda_1}} & \frac{\rho}{\sqrt{\lambda_1}} & 0 & \ldots & 0 \\ \frac{\alpha_2}{\sqrt{\lambda_2}} & 0 & \frac{\rho}{\sqrt{\lambda_2}} & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\alpha_{n-1}}{\sqrt{\lambda_{n-1}}} & 0 & 0 & \ldots & \frac{\rho}{\sqrt{\lambda_{n-1}}} \\ \frac{\alpha_n}{\sqrt{\lambda_n}} & -\frac{\alpha_1\rho}{\alpha_n\sqrt{\lambda_n}} & 0 & \ldots & -\frac{\alpha_{n-1}\rho}{\alpha_n\sqrt{\lambda_n}} \end{vmatrix} = \frac{(-1)^{n+1}\rho^{n-1}}{\alpha_n\sqrt{\lambda_1\ldots\lambda_n}}.$$

---

*) $\sqrt{\Lambda}$ представляет собой диагональную матрицу с элементами $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$.

А так как $u^T\Lambda u = \rho\alpha^T\alpha\rho = \rho^2$, то наш интеграл принимает вид

$$F_2(y) = \int\limits_{\rho^2 < y} p(\rho^2)\rho^{n-1}\,d\rho \int\limits_{\alpha_1^2+\ldots+\alpha_{n-1}^2<1}\ldots\int \frac{d\alpha_1\ldots d\alpha_{n-1}}{|\alpha_n|\sqrt{\lambda_1\ldots\lambda_n}}.$$

Интеграл по переменным $\alpha_1, \ldots, \alpha_{n-1}$ представляет собой постоянную величину c. Поэтому

$$F_2(y) = c\int\limits_0^{\sqrt{y}} p(\rho^2)\rho^{n-1}\,d\rho.$$

Дифференцируя эту формулу по y, находим плотность величины Y:

$$f_2(y) = \frac{c}{2}y^{\frac{n}{2}-1}p(y)\,1(y).$$

Постоянная c определяется из условия $F_2(\infty) = 1$.

В частном случае нормально распределенного вектора X

$$p(x^TCx) = \sqrt{\frac{|C|}{(2\pi)^n}}\,e^{-\frac{1}{2}x^TCx}$$

и полученная формула принимает вид

$$f_2(y) = p_n(y) = \frac{1(y)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{1}{2}y}. \qquad (11)$$

Это распределение, часто встречающееся в математической статистике, называется χ^2-*распределением*, так как величина с таким распределением в статистике называется *хи-квадрат*. Натуральное число n в (11) называется *числом степеней свободы*. Сравнение (11) с (2.15) показывает, что χ^2-распределение представляет собой γ-распределение с параметрами $k = 1/2$, $\mu = (n/2) - 1$.

Пример 11. Распределение случайных величин X и Y определяется формулой

$$f_1(x, y) = x^{\frac{k}{2}-1}y^{\frac{l}{2}-1}q(x+y)\,1(x)\,1(y),$$

где $q(t)$ — произвольная функция (которая может содержать δ-функции). Найти распределение отношения $Z = Y/X$.

Подставив данное выражение $f_1(x, y)$ в (10), получим

$$f_2(z) = z^{\frac{l}{2}-1}\int\limits_0^{\infty} x^{\frac{k+l}{2}-1}q(x(1+z))\,dx.$$

Замена переменных $u = x(1+z)$ дает

$$f_2(z) = z^{\frac{l}{2}-1}(1+z)^{-\frac{k+l}{2}} 1(z) \int_0^\infty u^{\frac{k+l}{2}-1} q(u)\, du.$$

Последний интеграл представляет собой постоянную величину. Его можно вычислить из условия нормировки (равенства интеграла от плотности единице). В результате получаем

$$f_2(z) = \frac{\Gamma\left(\frac{k+l}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{l}{2}\right)} z^{\frac{l}{2}-1}(1+z)^{-\frac{k+l}{2}} 1(z). \qquad (12)$$

Подчеркнем, что это распределение не зависит от вида функции $q(t)$. Таким образом, при любой функции $q(t)$ в выражении плотности $f_1(x, y)$ распределение отношения $Z = Y/X$ определяется плотностью (12).

Положив, в частности, $q(t) = ce^{-t/2}$ при соответствующем выборе постоянной c, получаем следующий результат: *плотность отношения двух независимых случайных величин X и Y, имеющих χ^2-распределения с k и l степенями свободы соответственно, определяется формулой* (12).

Пример 12. Распределение n-мерного случайного вектора X имеет шаровую симметрию, $f_1(x) = p(x^T x)$. Найти распределение случайной величины

$$T = \bar{X}\sqrt{\frac{n(n-1)}{X^T X - n\bar{X}^2}}, \qquad \bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k. \qquad (13)$$

По формуле (9) находим

$$F_2(t) = \int_{A_t} p(x^T x)\, dx,$$

где множество A_t определяется формулой

$$A_t = \left\{x\colon \bar{x}\sqrt{\frac{n(n-1)}{x^T x - n\bar{x}^2}} < t, \quad \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k\right\}.$$

Выполним такое ортогональное преобразование вектора x, при котором величина $\bar{x}\sqrt{n}$ переходит в первую координату y_1 вектора y. Для этого достаточно, чтобы все элементы первой строки матрицы преобразования A были равны $n^{-1/2}$. Тогда, имея в виду, что $A^T = A^{-1}$, $|A| = 1$, получим

$$F_2(t) = \int\limits_{y_1\sqrt{\frac{n-1}{y^T y - y_1^2}} < t} p(y^T y)\, dy.$$

Положив $y = \rho\alpha$, где ρ и α имеют тот же смысл, что и в примере 10, и имея в виду, что якобиан преобразования в данном случае равен

$(-1)^{n+1}\rho^{n-1}/\alpha_n$, получим

$$F_2(t)=c\int\limits_{\alpha_1\sqrt{\frac{n-1}{1-\alpha_1^2}}<t}\dots\int\frac{d\alpha_1\dots d\alpha_{n-1}}{|\alpha_n|},$$

где для краткости положено

$$c=\int\limits_0^\infty \rho^{n-1}p(\rho)\,d\rho.$$

Выразив в явной форме пределы интегрирования по переменным $\alpha_1, \dots, \alpha_{n-1}$, получим

$$F_2(t)=\int\limits_{\alpha_1\sqrt{\frac{n-1}{1-\alpha_1^2}}<t} dx_1 \int\limits_{\alpha_2^2+\dots+\alpha_{n-1}^2<1-\alpha_1^2}\dots\int\frac{d\alpha_2\dots d\alpha_{n-1}}{\sqrt{1-\alpha_1^2-\dots-\alpha_{n-1}^2}}.$$

Интегрирование по $\alpha_2, \dots, \alpha_{n-1}$ производится по $(n-2)$-мерному шару радиуса $\sqrt{1-\alpha_1^2}$, объем которого пропорционален $(1-\alpha_1^2)^{(n-2)/2}$. При этом подынтегральная функция в центре шара равна $(1-\alpha_1^2)^{-1/2}$. Поэтому интуитивно ясно, что интеграл по переменным $\alpha_2, \dots, \alpha_{n-1}$ пропорционален $(1-\alpha_1^2)^{(n-3)/2}$. Чтобы строго доказать это, достаточно ввести новые переменные интегрирования $\beta_2=\alpha_2/\sqrt{1-\alpha_1^2}, \dots, \beta_{n-1}=\alpha_{n-1}/\sqrt{1-\alpha_1^2}$. Тогда, объединив постоянный интеграл по переменным $\beta_2, \dots, \beta_{n-1}$ с постоянным множителем c и обозначив новую постоянную буквой a, получим

$$F_2(t)=a\int\limits_{\alpha_1\sqrt{\frac{n-1}{1-\alpha_1^2}}<t}(1-\alpha_1^2)^{\frac{n-3}{2}}\,d\alpha_1.$$

Наконец, вводя новую переменную интегрирования $\tau=\alpha_1\sqrt{(n-1)/(1-\alpha_1^2)}$, приведем полученную формулу к виду

$$F_2(t)=a\int\limits_{-\infty}^{t}\left(1+\frac{\tau^2}{n-1}\right)^{-n/2}d\tau.$$

Дифференцируя эту формулу, находим плотность величины T:

$$f_2(t)=a\left(1+\frac{t^2}{n-1}\right)^{-n/2}.$$

Постоянная a легко определяется из условия $F_2(\infty)=1$. Для вычисления интеграла достаточно сделать замену переменных $\operatorname{tg}\varphi=t/\sqrt{n-1}$ и принять во внимание, что $\Gamma(1/2)=\sqrt{\pi}$. Тогда

получим

$$f_2(t) = s_{n-1}(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{t^2}{n-1}\right)^{-n/2}. \qquad (14)$$

Это распределение, часто встречающееся в математической статистике, называется *распределением Стьюдента* или, короче, *T-распределением*. Величина T, определяемая формулой (13), называется *стьюдентовым отношением* *). Число $k = n-1$ называется *числом степеней свободы* [19, 115].

Полезно отметить, что распределение стьюдентова отношения T не зависит от функции $p(u)$. Для любого случайного вектора X, распределение которого обладает шаровой симметрией, случайная величина T имеет T-распределение (14).

Пример 13. В условиях примера 12 найти совместное распределение случайных величин $U = \bar{X}$ и $S = X^T X - n\bar{X}^2$.

По формуле (9) находим

$$F_2(u, s) = \int\limits_{\substack{\bar{x} < u \\ x^T x - n\bar{x}^2 < s}} p(x^T x)\,dx.$$

Выполнив ортогональное преобразование предыдущего примера, получим

$$F_2(u, s) = \int\limits_{-\infty}^{u\sqrt{n}} dy_1 \int\limits_{y_2^2 + \ldots + y_n^2 < s} \!\!\ldots\!\! \int p(y^T y)\,dy_2 \ldots dy_n.$$

После замены переменных $y_2 = \rho\beta_2, \ldots, y_n = \rho\beta_n$, $\beta_2^2 + \ldots + \beta_n^2 = 1$, так же как в предыдущем примере, получим

$$F_2(u, s) = c \int\limits_{-\infty}^{u\sqrt{n}} dy_1 \int\limits_0^{\sqrt{s}} \rho^{n-2} p\left(y_1^2 + \rho^2\right) d\rho,$$

где c определяется из условия $F_2(\infty, \infty) = 1$.

Доказать, что при любой функции $p(v)$ случайные величины U и S не коррелированы и что они независимы тогда и только тогда, когда $p(v) = (2\pi D)^{-n/2} e^{-v/2D}$.

Пример 14. Показать, что если совместное распределение n-мерного вектора X и m-мерного вектора Y обладает шаровой симметрией в $(n+m)$-мерном пространстве, то случайная величина

$$Z = \sqrt{\frac{nm(n+m-2)}{(n+m)(X^T X - n\bar{X}^2 + Y^T Y - m\bar{Y})^2}}\,(\bar{X} - \bar{Y}), \qquad (15)$$

*) Стьюдент (Student) — псевдоним английского ученого Госсета (W. S. Gosset).

где

$$\bar{X}=\frac{1}{n}\sum_{p=1}^{n}X_p, \qquad \bar{Y}=\frac{1}{m}\sum_{q=1}^{m}Y_q,$$

имеет T-распределение с $n+m-2$ степенями свободы:

$$s_k(z)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{z^2}{k}\right)^{-\frac{k+1}{2}}, \quad k=n+m-2. \tag{16}$$

Пример 15. Пусть X — случайная $n\times m$-матрица, $m<n$, $\bar{X}$ — матрица-строка, элементами которой служат средние арифметические элементов соответствующих столбцов матрицы X:

$$\bar{X}_p=\frac{1}{n}\sum_{q=1}^{n}X_{qp} \qquad (p=1, \ldots, m).$$

Случайная величина

$$T=\sqrt{n(n-1)\bar{X}S^{-1}\bar{X}^T}, \qquad S=X^TX-n\bar{X}^T\bar{X}, \tag{17}$$

называется *обобщенным стьюдентовым отношением*. Для любой случайной матрицы X, плотность которой определяется формулой

$$f_1(x)=p(\operatorname{tr} xCx^T), \tag{18}$$

где $p(u)$ — любая функция (конечно, удовлетворяющая условию, чтобы функция $f_1(x)$ была плотностью, возможно содержащей δ-функции), C — положительно определенная симметричная матрица, а $\operatorname{tr} A$ — след матрицы A, случайная величина T имеет плотность

$$s_{n-1,m}(t)=\frac{2\Gamma\left(\frac{n}{2}\right)}{(n-1)^{m/2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-m}{2}\right)}t^{m-1}\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}1(t). \tag{19}$$

В частном случае при $m=1$ это распределение совпадает с распределением абсолютной величины стьюдентова отношения (13).

Пример 16. В условиях предыдущего примера найти совместное распределение случайного вектора (матрицы-строки) $\bar{X}$ и случайной матрицы S. Показать, что они не коррелированы при любой функции $p(u)$ и что они независимы тогда и только тогда, когда $p(u)=\sqrt{|C|^n/(2\pi)^{nm}}\,e^{-u/2}$.

Пример 17. Пусть X_1 и X_2 — случайные матрицы $n_1\times m$ и $n_2\times m$ соответственно, $\bar{X}_1$ и $\bar{X}_2$ — матрицы-строки, элементами которых служат средние арифметические элементов соответствующих столбцов матриц X_1 и X_2, $S_k=X_k^TX_k-n\bar{X}_k^T\bar{X}_k$ $(k=1, 2)$, $H=S_1+S_2$. Показать, что если распределение блочной $(n_1+n_2)\times m$-матрицы X, $X^T=\left[X_1^TX_2^T\right]$,

определяется формулой (18), то случайная величина

$$U=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}(\bar{X}_1-\bar{X}_2)H^{-1}(\bar{X}_1^T-\bar{X}_2^T)} \qquad (20)$$

имеет плотность (19), где $n=n_1+n_2-1$.

Пример 18. Если в условиях примера 15 $p(u)=\sqrt{|C|^n/(2\pi)^{nm}}\,e^{-u/2}$, а $C=K^{-1}$, то плотность матрицы S определяется формулой

$$w_{nm}(s)=c_{nm}(2^m|K|)^{-\frac{n-1}{2}}|s|^{\frac{n-m}{2}-1}e^{-\frac{1}{2}\operatorname{tr}K^{-1}s} \qquad (21)$$

в области положительно определенных матриц s пространства квадратных матриц порядка m и равна нулю вне этой области. Через $|K|$ и $|s|$ в формуле (21) обозначены соответственно определители матриц K и s, а

$$c_{nm}=\frac{1}{\pi^{\frac{m(m-1)}{4}}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)\ldots\Gamma\left(\frac{n-m}{2}\right)}.$$

Распределение, определяемое плотностью (21), называется *распределением Уишарта*. Оно играет большую роль в математической статистике [104, 105].

Пример 19. Если в условиях примера 15 $S'=X^TX$, то отношение определителей матриц S и S', $V=|S|/|S'|$, имеет плотность

$$f_2(v)=\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-m}{2}\right)\Gamma\left(\frac{m}{2}\right)}v^{\frac{n-m}{2}-1}(1-v)^{\frac{m}{2}-1}1(v)1(1-v). \qquad (22)$$

Это β-распределение с параметрами $p=(n-m)/2$ и $q=m/2$ (пример 2.9).

Пример 20. Функция распределения $F_1(x)$ скалярной случайной величины X непрерывна. Найти распределение случайной величины $Y=F_1(X)$.

При любом $y\in(0,1]$ событие $Y=F_1(X)<y$ происходит тогда и только тогда, когда $X<x=F_1^{-1}(y)$. Поэтому

$$F_2(y)=F_1(F_1^{-1}(y))=y \text{ при } y\in(0,1].$$

При $y\leqslant 0$ событие $Y=F_1(X)<y$ невозможно, а при $y>1$ — достоверно. Поэтому $F_2(y)=0$ при $y\leqslant 0$ и $F_2(y)=1$ при $y>1$. Таким образом, при любой непрерывной функции распределения $F_1(x)$ случайная величина $Y=F_1(X)$ распределена равномерно на интервале $[0,1]$.

Предлагаем читателю самостоятельно найти распределение случайной величины $Y=F_1(X)$ в случае, когда функция распределения $F_1(x)$ имеет разрывы. Найти плотность величины X в этом случае.

2.3. Приведение случайного вектора к вектору с независимыми координатами. Рассмотрим n-мерный случайный вектор $X=\{X_1,\ldots,X_n\}$, у которого непрерывны

функция распределения $F_1(x_1)$ первой координаты и условные функции распределения $F_2(x_2 \mid x_1), \ldots$ $\ldots, F_n(x_n \mid x_1, \ldots, x_{n-1})$. Этому условию, в частности, удовлетворяет любой случайный вектор, плотность которого не содержит δ-функций. Во многих задачах теории вероятностей и ее приложений оказывается полезным преобразовать случайный вектор так, чтобы в результате получить вектор с независимыми координатами.

На основании результатов п. 4.4.5 эта задача легко решается для нормально распределенных векторов применением канонических разложений, в частности разложения по собственным векторам.

► Для преобразования случайного вектора с произвольным распределением, удовлетворяющим поставленным условиям, можно воспользоваться результатом примера 20. Введем случайные величины

$$Y_1 = F_1(X_1), \quad Y_2 = F_2(X_2 \mid X_1), \ldots$$
$$\ldots, Y_n = F_n(X_n \mid X_1, \ldots, X_{n-1}). \quad (23)$$

Так как $F_1(x_1), F_2(x_2 \mid x_1), \ldots, F_n(x_n \mid x_1, \ldots, x_{n-1})$, рассматриваемые как функции первого аргумента, представляют собой непрерывные функции распределения, то по доказанному в примере 20 случайная величина Y_1 и все последующие случайные величины Y_k, независимо от значений $y_1, \ldots, y_{k-1} \in [0, 1]$ предыдущих величин $Y_1, \ldots$ $\ldots, Y_{k-1}$ ($k = 2, \ldots, n$), распределены равномерно на интервале $[0, 1]$. Таким образом, условное распределение каждой величины Y_k не зависит от значений $y_1, \ldots, y_{k-1}$ предшествующих величин. Следовательно, совместная плотность величин $Y_1, \ldots, Y_n$ равна произведению их плотностей. Отсюда по доказанному в п. 4.2.3 вытекает, что случайные величины $Y_1, \ldots, Y_n$ независимы. ◄

Итак, формулы (23) решают поставленную задачу — определяют такое преобразование случайного вектора X, в результате которого получается вектор Y с независимыми и вдобавок равномерно распределенными на интервале $[0, 1]$ координатами (отсюда следует, что вектор Y равномерно распределен в n-мерном кубе со сторонами $[0, 1]$ по всем осям). Обратное преобразование, определяемое рекуррентными формулами

$$X_1 = F_1^{-1}(Y_1), \quad X_2 = F_2^{-1}(Y_2 \mid X_1), \ldots$$
$$\ldots, X_n = F_n^{-1}(Y_n \mid X_1, \ldots, X_{n-1}), \quad (24)$$

дает выражение координат вектора X через независимые

равномерно распределенные на интервале [0, 1] случайные величины $Y_1, \ldots, Y_n$.

Теперь мы можем найти такое преобразование случайного вектора X, в результате которого получается случайный вектор Z с независимыми координатами, каждая из которых имеет наперед заданную функцию распределения.

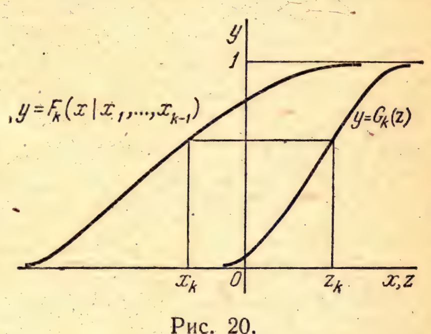


Рис. 20.

▶ Пусть $G_1(z_1), \ldots, G_n(z_n)$ — заданные строго возрастающие функции распределения. Определим координаты вектора Z уравнениями

$$G_k(Z_k) = Y_k \quad (k = 1, \ldots, n), \qquad (25)$$

где $Y_1, \ldots, Y_n$ — случайные величины, определяемые формулами (23). Так как функции $G_1, \ldots, G_n$ по условию строго возрастающие, то формулы (25) устанавливают взаимно однозначное соответствие между величинами $Y_1, \ldots, Y_n$, с одной стороны, и величинами $Z_1, \ldots, Z_n$, с другой. Очевидно, что случайные величины $Z_1, \ldots, Z_n$ независимы как функции независимых случайных величин $Y_1, \ldots, Y_n$ (п. 4.2.5), и функцией распределения величины Z_k служит $G_k(z_k)$. Действительно,

$$P(Z_k < z_k) = P(G_k(Z_k) < G_k(z_k)) = \\ = P(Y_k < G_k(z_k)) = G_k(z_k). \blacktriangleleft$$

Таким образом, формулы (23) и (25) определяют преобразование случайного вектора X, в результате которого получается вектор Z с независимыми координатами $Z_1, \ldots, Z_n$, имеющими заданные функции распределения $G_1(z_1), \ldots, G_n(z_n)$. Обратное преобразование, определяемое рекуррентными формулами

$$X_1 = F_1^{-1}(G_1(Z_1)), \quad X_2 = F_2^{-1}(G_2(Z_2) \mid X_1), \ \ldots \\ \ldots, X_n = F_n^{-1}(G_n(Z_n) \mid X_1, \ldots, X_{n-1}), \qquad (26)$$

выражает вектор X через вектор Z с независимыми координатами, имеющими заданные функции распределения. На рис. 20 показано соответствие между значениями x_k и z_k случайных величин X_k и Z_k.

Заметим, что в случае, когда функция распределения одной из величин $X_1, \ldots, X_n$ сохраняет постоянное значение на каком-нибудь интервале, взаимная однозначность соответствия между векторами X и Y и векторами X и Z нарушается. В этом случае формулы (24) и (25) не определяют значения x вектора X при некоторых значениях y вектора Y. Однако это не существенно, потому что каждый интервал постоянства функции распределения представляет собой множество нулевой вероятности, с которым можно не считаться. Впрочем, можно принять, что в данном случае формулы (24) и (25) для соответствующих y определяют весь интервал постоянства функции распределения.

§ 3. Плотность функции случайного аргумента

3.1. Метод сравнения вероятностей. В § 2 мы научились находить функцию распределения функции случайного аргумента, а затем дифференцированием функции распределения определять плотность, если, конечно, она существует. Однако часто целесообразно непосредственно находить плотность функции случайной величины по данной плотности величины-аргумента. При этом на функцию $\varphi(x)$ придется наложить дополнительные ограничения, одной ее измеримости в общем случае недостаточно. Мы будем предполагать, что функция $\varphi(x)$ имеет кусочно непрерывные первые производные по всем координатам вектора x и не постоянна ни на каком множестве значений аргумента x, имеющем отличную от нуля вероятность.

▶ Согласно общему принципу п. 2.1 вероятность попадания случайной величины $Y=\varphi(X)$ на любое множество B равна вероятности попадания величины X на прообраз множества B, $A_B=\{x\colon \varphi(x)\in B\}$. Выразив эту вероятность через плотность $f_1(x)$ случайной величины X, получим

$$P(Y\in B)=P(\varphi(X)\in B)=P(X\in A_B)=\int_{A_B} f_1(x)\,dx. \quad (27)$$

С другой стороны,

$$P(Y\in B)=\int_B f_2(y)\,dy, \quad (28)$$

где $f_2(y)$ — неизвестная плотность случайной величины Y. Сравнение формул (27) и (28) показывает, что для опре-

деления $f_2(y)$ достаточно заменой переменной преобразовать интеграл (27) в интеграл, распространенный на область B.

Предположим сначала, что векторы X и $Y=\varphi(X)$ имеют одну и ту же размерность n и уравнение $y=\varphi(x)$ при любом y имеет единственное решение x в области возможных значений случайной величины X (в области, где плотность $f_1(x)$ отлична от нуля). Для этого необходимо и достаточно, чтобы якобиан координат вектора

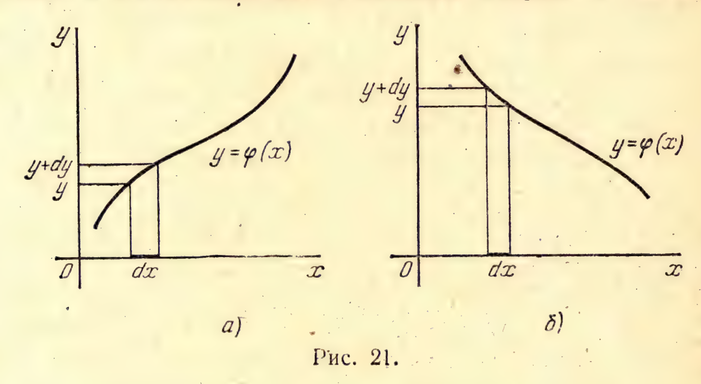


Рис. 21.

$\varphi(x)$ по координатам вектора x сохранял знак в области, где $f_1(x)>0$, и обращался в нуль только в изолированных точках этой области. На рис. 21 дана иллюстрация для случая скалярных X и Y. Сделав в формуле (27) замену переменных $y=\varphi(x)$, $x=\varphi^{-1}(y)$ и приняв во внимание, что область интегрирования A_B перейдет при этом в B, будем иметь

$$P(Y\in B)=\int_B f_1(\varphi^{-1}(y))\,|J(y)|\,dy, \tag{29}$$

где $J(y)$ — якобиан координат вектора $x=\varphi^{-1}(y)$ по координатам вектора y:

$$J(y)=\frac{\partial(\varphi_1^{-1},\ \ldots,\ \varphi_n^{-1})}{\partial(y_1,\ \ldots,\ y_n)}=\begin{vmatrix}\dfrac{\partial\varphi_1^{-1}}{\partial y_1} & \ldots & \dfrac{\partial\varphi_1^{-1}}{\partial y_n}\\ \cdot & \cdot\ \cdot\ \cdot & \cdot\\ \dfrac{\partial\varphi_n^{-1}}{\partial y_1} & \ldots & \dfrac{\partial\varphi_n^{-1}}{\partial y_n}\end{vmatrix} \tag{30}$$

(производная функции $\varphi^{-1}(y)$ в случае скалярных X и Y). Сравнив (29) с (28) и приняв во внимание, что обе эти формулы справедливы для любой области B, полу-

чаем следующее выражение для плотности случайной величины $Y=\varphi(X)$:

$$f_2(y)=f_1(\varphi^{-1}(y))\,|J(y)|. \qquad (31)$$

Предположим теперь, что размерность m вектора $Y=\varphi(X)$ меньше размерности n вектора X, $m<n$, и уравнение $y=\varphi(x)$ при всех y имеет единственное решение относительно каких-нибудь m координат вектора x в области, где $f_1(x)>0$. Обозначим через x' вектор, образованный этими m координатами, через x'' вектор, образованный $n-m$ оставшимися координатами вектора x. Тогда наше уравнение перепишется в виде $y=\varphi(x', x'')$. Пусть $x'=\varphi^{-1}(y, x'')$ — решение этого уравнения относительно x'. Сделав в (27) замену переменных $x'=\varphi^{-1}(y, x'')$, получим совершенно так же, как и в предыдущем случае,

$$P(Y\in B)=\int_B dy \int_{-\infty}^{\infty} f_1(\varphi^{-1}(y, x''),\ x'')\,|J(y, x'')|\,dx'', \qquad (32)$$

где $J(y, x'')$ — якобиан координат вектора $x'=\varphi^{-1}(y, x'')$ по координатам вектора y:

$$J(y, x'')=\frac{\partial(\varphi_1^{-1}, \ldots, \varphi_m^{-1})}{\partial(y_1, \ldots, y_m)}. \qquad (33)$$

Сравнив (32) с (28), находим плотность случайной величины Y:

$$f_2(y)=\int_{-\infty}^{\infty} f_1(\varphi^{-1}(y, u), u)\,|J(y, u)|\,du. \blacktriangleleft \qquad (34)$$

Эту формулу можно получить также другим путем. Дополнив вектор Y до n-мерного вектора, добавив к нему координаты $Y_{m+1}=X_{m+1}, \ldots, Y_n=X_n$, приведем задачу к предыдущему случаю. Определив по формуле (31) совместную плотность векторов Y и $X''=\{X_{m+1}, \ldots, X_n\}$, можем после этого найти плотность вектора Y по формуле (4.9) интегрированием по x'' по области всех возможных значений случайного вектора X''.

Если векторы X и $Y=\varphi(X)$ имеют одинаковую размерность n, а уравнение $y=\varphi(x)$ при каждом y имеет конечное или счетное множество решений относительно x в области возможных значений вектора X, то область интегрирования A_B следует разбить на части, в каждой из которых содержится не больше одного решения этого

уравнения, а затем выполнить соответствующую замену переменных в каждом из полученных интегралов. Предоставляем читателю самостоятельно найти плотность случайной величины $Y=\varphi(X)$ таким путем. Мы решим здесь эту задачу более удобным для данного случая методом сравнения элементов вероятности.

3.2. Метод сравнения элементов вероятности. Метод сравнения элементов вероятности основан на выборе

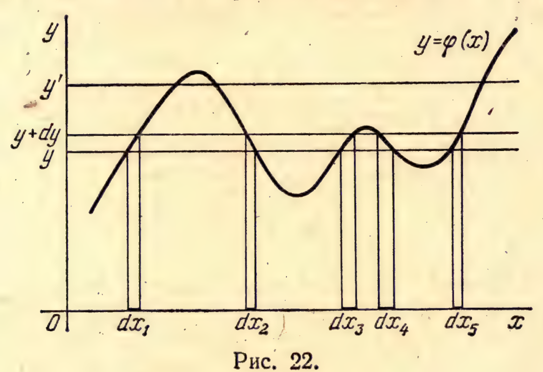


Рис. 22.

в качестве множества B бесконечно малого элемента dy пространства значений случайной величины Y и нахождении вероятности попадания величины $Y=\varphi(X)$ в этот элемент. Множитель при dy будет в этом случае искомой плотностью случайной величины $Y=\varphi(X)$.

▶ Предположим, что векторы X и Y имеют одинаковую размерность n и уравнение $y=\varphi(x)$ при каждом y имеет множество решений $\varphi_i^{-1}(y)$, $i\in I(y)$, в области возможных значений величины X (т. е. в области, где $f_1(x)\neq 0$). На рис. 22 дана иллюстрация для случая скалярных величин X и Y (на рис. 22, как мы видим, $I(y)=\{1,\ 2,\ 3,\ 4,\ 5\}$, а $I(y')=\{1,\ 2,\ 5\}$). В этом случае элемент вероятности $f_2(y)\,dy$ равен сумме элементов вероятности, соответствующих всем элементам dx_i, $i\in I(y)$, пространства значений величины X, объединение которых представляет собой прообраз элемента dy (рис. 22):

$$f_2(y)\,dy=\sum_{i\in I(y)} f_1(x_i)\,dx_i.$$

Подставив сюда выражения x_i через y, $x_i=\varphi_i^{-1}(y)$, и выражения элементов dx_i через dy, $dx_i=|J_i(y)|\,dy$, где $J_i(y)$ — якобиан координат вектора $x_i=\varphi_i^{-1}(y)$ по коор-

динатам вектора y (производная функции $\varphi_i^{-1}(y)$ в случае скалярных X и Y), получим после сокращения на dy

$$f_2(y) = \sum_{i \in I(y)} f_1(\varphi_i^{-1}(y)) \, |J_i(y)|. \blacktriangleleft \qquad (35)$$

Метод сравнения элементов вероятности представляет собой сокращенную форму следующей последовательности вполне строгих операций: 1) вычисление вероятности попадания величины $Y = \varphi(X)$ в бесконечно малый объем Δy ее пространства значений, содержащий данную точку y, с помощью интегральной теоремы о среднем значении; 2) вычисление таким же путем вероятности попадания величины X в объединение элементарных объемов Δx_i, $i \in I(y)$, ее пространства значений, представляющее собой прообраз Δy,

$$\bigcup_{i \in I(y)} \Delta x_i = \varphi^{-1}(\Delta y);$$

3) приравнивание полученных таким путем вероятностей; 4) нахождение выражений элементарных объемов Δx_i через Δy; 5) деление полученного равенства на Δy; 6) переход к пределу, когда элементарный объем Δy стягивается в точку y, с попутным доказательством существования предела.

▶ В случае, когда вектор $Y = \varphi(X)$ имеет размерность $m < n$, а уравнение $y = \varphi(x)$ не имеет единственного решения относительно m координат вектора x ни при каком выборе этих координат, следует дополнить вектор Y до n-мерного добавлением координат $Y_{m+1} = X_{m+1}, \ldots$ $\ldots, Y_n = X_n$, вычислить плотность полученного n-мерного вектора по формуле (35), а затем интегрированием по вектору $u = \{y_{m+1}, \ldots, y_n\}$ найти плотность величины Y. В результате получим формулу

$$f_2(y) = \int_{-\infty}^{\infty} \sum_{i \in I(y, u)} f_1(\varphi_i^{-1}(y, u), u) \, |J_i(y, u)| \, du, \qquad (36)$$

где $\varphi_i^{-1}(y, u)$, $i \in I(y, u)$, — решения уравнения $y = \varphi(x)$ относительно вектора $x' = \{x_1, \ldots, x_m\}$ в области возможных значений величины X, а $J_i(y, u)$ — якобиан координат вектора $x_i' = \varphi_i^{-1}(y, u)$ по координатам вектора y. ◀

▶ Рассмотрим, наконец, случай, когда вектор $Y = \varphi(X)$ имеет размерность $m > n$. В этом случае вектор Y можно

представить как совокупность n-мерного вектора $Y'=\varphi_1(X)$ и $(m-n)$-мерного вектора $Y''=\varphi_2(X)$. При этом разбиении вектора Y (его замене проекциями на два взаимно дополнительных подпространства) следует включить в Y' такие координаты вектора Y, чтобы уравнение $y'=\varphi_1(x)$ имело единственное решение x в области возможных значений величины X, если, конечно, это возможно при каком-нибудь выборе n из m координат вектора Y. Если это уравнение имеет единственное решение $x=\varphi_1^{-1}(y')$, то плотность случайного вектора Y' можно вычислить по формуле (31). Что касается вектора Y'', то, подставив в $Y''=\varphi_2(X)$ выражение $X=\varphi_1^{-1}(Y')$, получим $Y''=\varphi_2(\varphi_1^{-1}(Y'))$. Таким образом, Y'' представляет собой вполне определенную функцию случайной величины Y', и, следовательно, условная плотность Y'' при данном значении y' величины Y' представляет собой δ-функцию $\delta(y''-\varphi_2(\varphi_1^{-1}(y')))$. На основании теоремы умножения плотностей (4.23) плотность составного случайного вектора Y выразится в этом случае формулой

$$f_2(y)=f_1(\varphi_1^{-1}(y'))\,|J(y')|\,\delta(y''-\varphi_2(\varphi_1^{-1}(y'))). \blacktriangleleft \quad (37)$$

Если уравнение $y'=\varphi_1(x)$ при любом выборе n из m координат вектора y имеет множество решений $\varphi_{1i}^{-1}(y')$, $i\in I(y')$, в области возможных значений величины X, то вектор Y'' при данном значении y' вектора Y' имеет дискретное множество возможных значений $\varphi_2(\varphi_{1i}^{-1}(y'))$, $i\in I(y')$, вероятности которых определяются формулой

$$p_i=\frac{f_1(x_i)\,dx_i}{\sum\limits_{k\in I(y')} f_1(x_k)\,dx_k}=\frac{f_1(\varphi_{1i}^{-1}(y'))\,|J_i(y')|}{\sum\limits_{k\in I(y')} f_1(\varphi_{1k}^{-1}(y'))\,|J_k(y')|}.$$

Следовательно, условная плотность вектора Y'' при данном значении y' вектора Y' согласно (2.24) равна

$$\frac{\sum\limits_{i\in I(y')} f_1(\varphi_{1i}^{-1}(y'))\,|J_i(y')|\,\delta(y''-\varphi_2(\varphi_{1i}^{-1}(y')))}{\sum\limits_{i\in I(y')} f_1(\varphi_{1i}^{-1}(y'))\,|J_i(y')|}.$$

Умножив эту условную плотность величины Y'' на плотность величины Y', вычисленную по формуле (35), найдем плотность случайного вектора Y:

$$f_2(y)=\sum_{i\in I(y')} f_1(\varphi_{1i}^{-1}(y'))\,|J_i(y')|\,\delta(y''-\varphi_2(\varphi_{1i}^{-1}(y'))). \quad (38)$$

Пример 21. Найти распределение s-й степени случайной величины X, $Y=X^s$ (s — натуральное число).

При нечетном s уравнение $y=x^s$ имеет единственное действительное решение $x=y^{1/s}$ при любом y. Поэтому плотность величины $Y=X^s$ можно найти по формуле (31):

$$f_2(y)=\frac{1}{s}f_1(y^{1/s})\,|y|^{\frac{1}{s}-1}.$$

При четном s уравнение $y=x^s$ имеет два действительных решения $x=\pm y^{1/s}$ при $y\geqslant 0$ и ни одного при $y<0$. Поэтому для нахождения плотности величины $Y=X^s$ придется применить формулу (35). В результате получим

$$f_2(y)=\frac{1}{s}\left[f_1\left(y^{\frac{1}{s}}\right)+f_1\left(-y^{\frac{1}{s}}\right)\right]y^{\frac{1}{s}-1}\,1(y).$$

В частном случае величины X с плотностью

$$f_1(x)=\frac{1(x)}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}x^{n-1}e^{-\frac{1}{2}x^2}$$

и $Y=X^2$ полученная формула дает

$$f_2(y)=\frac{1(y)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{1}{2}y}.$$

Таким образом, в этом случае величина $Y=\chi^2$ имеет χ^2-распределение (11). Поэтому распределение величины X с такой плотностью $f_1(x)$ называется χ-распределением.

Рис. 23.

Пример 22. Найти плотность случайной величины $Y=\cos X$ при равномерном распределении величины X в интервале $(-\pi, \pi)$.

В данном случае при каждом значении y, $|y|<1$, обратная функция имеет две ветви: $\varphi_1^{-1}(y)=\arccos y$, $\varphi_2^{-1}(y)=-\arccos y$ (рис. 23). Формула (35) дает

$$f_2(y)=\frac{1(1-|y|)}{\pi\sqrt{1-y^2}}.$$

Распределение вероятностей, определяемое этой формулой, называется *распределением арксинуса*. Оно характерно для ошибок угломерного прибора с эксцентрично закрепленным по отношению к оси вращения визирного устройства лимбом (диском, на котором нанесена шкала).

Пример 23. Найти плотность функции

$$Y=X^2-kX$$

случайной величины X, равномерно распределенной в интервале $(0, a)$, $k < a$.

В данном случае обратная функция имеет две ветви

$$\varphi_1^{-1}(y) = \frac{k}{2} + \sqrt{\frac{k^2}{4} + y}, \quad \varphi_2^{-1}(y) = \frac{k}{2} - \sqrt{\frac{k^2}{4} + y}$$

при $-k^2/4 < y < 0$ и одну только первую ветвь при $0 < y < a^2 - ka$ (рис. 24). Поэтому в правой части формулы (35) будет два слагаемых при $-k^2/4 < y < 0$ и одно первое слагаемое при $0 < y < a^2 - ka$. В результате получим

$$f_2(y) = \begin{cases} \dfrac{2}{a\sqrt{k^2+4y}} & \text{при} \quad -\dfrac{k^2}{4} < y < 0, \\ \dfrac{1}{a\sqrt{k^2+4y}} & \text{при} \quad 0 < y < a^2 - ka. \end{cases}$$

Вне интервала $(-k^2/4,\ a^2 - ka)$ $f_2(y) = 0$.

Пример 24. В условиях примера 11 найти распределение случайной величины

$$F = \frac{Y/l}{X/k} = \frac{k}{l} Z.$$

Так как уравнение $f = kz/l$ при любом f имеет единственное решение $z = lf/k$, то плотность $g(f)$ величины F можно определить по формуле (31). Подставив в нее вместо $f_1(x)$ выражение (12) плотности $f_2(z)$ величины Z, получим

$$g(f) = f_{lk}(f) = \left(\frac{l}{k}\right)^{\frac{l}{2}} \frac{\Gamma\left(\frac{k+l}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{l}{2}\right)} f^{\frac{l}{2}-1} \left(1 + \frac{l}{k} f\right)^{-\frac{k+l}{2}} 1(f). \quad (39)$$

Это распределение тоже часто встречается в математической статистике и называется *F-распределением с l и k степенями свободы.* Распределение величины $U = \ln F/2$ в частном случае, когда X и Y — независимые случайные величины, имеющие χ^2-распределение соответственно с k и l степенями свободы, было впервые найдено Фишером [112, 113]. Распределение величины F при тех же условиях получено Снедекором [99].

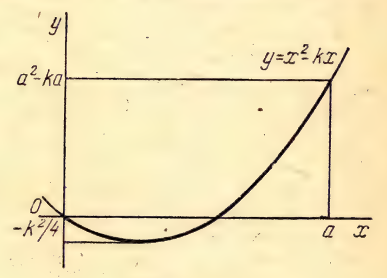


Рис. 24.

При $q(t) = ce^{-t/2}$ при соответствующем выборе c из полученного результата следует, что *отношение двух независимых величин с χ^2-распределением, деленных на соответствующие числа степеней свободы, имеет F-распределение с этими числами степеней свободы.* Этим и объясняется широкое распространение F-распределения.

Интересно отметить, что предельные случаи F-распределения при $k=\infty$ и при $l=\infty$ представляют собой γ-распределения величин F и F^{-1} соответственно:

$$f_{l,\,\infty}(f)=\left(\frac{l}{2}\right)^{l/2}\frac{f^{l/2-1}}{\Gamma(l/2)}e^{-lf/2},$$

$$f_{\infty,\,k}(f)=\left(\frac{k}{2}\right)^{k/2}\frac{f^{-k/2-1}}{\Gamma(k/2)}e^{-k/2f}.$$

В предельном случае при $k=l=\infty$ $f_{\infty,\,\infty}(f)=\delta(f-1)$.

Пример 25. В условиях примера 15 случайная величина

$$F=\frac{n-m}{m(n-1)}T^2=\frac{n(n-m)}{m}\overline{X}S^{-1}\overline{X}^T$$

независимо от вида функции $p(u)$ имеет F-распределение с m и $n-m$ степенями свободы $f_{m,\,n-m}(f)$.

Пример 26. Если величина F имеет F-распределение $f_{lk}(f)$, то величина $V=(1+lF/k)^{-1}$ имеет β-распределение с параметрами $p=k/2$, $q=l/2$ (пример 2.9), а величина $W=F(k/l+F)^{-1}$ имеет β-распределение с параметрами $p=l/2$, $q=k/2$.

Отсюда следует, что если величина V имеет β-распределение с параметрами p, q, то величина $F=(k/l)V(1-V)^{-1}$ имеет F-распределение $f_{2p,\,2q}(f)$, а величина $G=k(1-V)/lV$ имеет F-распределение $f_{2q,\,2p}(g)$.

Пользуясь этими соотношениями и соотношением между F и T примера 25, можно вычислять функцию распределения и плотность для F-распределения и обобщенного T-распределения (19) с помощью программы BDTR вычисления функции распределения и плотности для β-распределения [94].

Пример 27. Найти распределение полярных координат точки на плоскости R, Θ, если известна плотность $f_1(x,\,y)$ декартовых координат X, Y.

В данном случае

$$R=\varphi_1(X,\,Y)=\sqrt{X^2+Y^2},\quad \Theta=\varphi_2(X,\,Y)=\operatorname{arctg}\frac{Y}{X},$$

причем угол Θ лежит в первой четверти при $X>0$, $Y>0$, во второй при $X<0$, $Y>0$, в третьей при $X<0$, $Y<0$ и в четвертой при $X>0$, $Y<0$. Уравнения $r=\sqrt{x^2+y^2}$, $\theta=\operatorname{arctg}(y/x)$ имеют единственное решение $x=r\cos\theta$, $y=r\sin\theta$ при любых $r\geqslant 0$, $\theta\in[0,\,2\pi)$. Поэтому

$$J(r,\,\theta)=\begin{vmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{vmatrix}=r$$

и формула (31) дает

$$f_2(r,\,\theta)=rf_1(r\cos\theta,\,r\sin\theta).$$

Отсюда интегрированием по θ или по r находим соответственно плотность $f_3(r)$ радиуса-вектора R случайной точки и плотность

$f_4(\theta)$ полярного угла Θ:

$$f_3(r) = r\int_0^{2\pi} f(r\cos\theta,\ r\sin\theta)\,d\theta,$$

$$f_4(\theta) = \int_0^{\infty} rf(r\cos\theta,\ r\sin\theta)\,dr.$$

В частном случае распределения с круговой симметрией $f(x, y) = h(x^2+y^2)$ и полученные формулы принимают вид

$$f_3(r) = 2\pi r h(r^2), \qquad f_4(\theta) = \int_0^{\infty} rh(r^2)\,dr.$$

Таким образом, в случае круговой симметрии распределения на плоскости полярный угол распределен равномерно в интервале $[0, 2\pi)$, и следовательно, $f_4(\theta) = 1/2\pi$ при $\theta \in [0, 2\pi)$. Впрочем, это ясно и из интуитивных соображений.

В случае нормального распределения с круговой симметрией $h(u) = (2\pi D)^{-1} e^{-u/2D}$ и формула для $f_3(r)$ принимает вид

$$f_3(r) = \frac{r}{D} e^{-r^2/2D}, \qquad r > 0.$$

Это распределение называется *распределением Рэлея*.

Аналогично в случае нормального распределения с шаровой симметрией в трехмерном пространстве получаем для радиуса-вектора R *распределение Максвел*

$$f_3(r) = \sqrt{2/\pi D^3}\, r^2 e^{-r^2/2D}, \qquad r > 0.$$

Если распределение на плоскости имеет круговую симметрию с центром, смещенным относительно начала координат, то, выбрав соответственно оси координат, будем иметь $f_1(x, y) = h((x-a)^2+y^2)$ и

$$f_3(r) = r\int_0^{2\pi} h(r^2 - 2ar\cos\theta + a^2)\,d\theta,$$

$$f_4(\theta) = \int_0^{\infty} rh(r^2 - 2ar\cos\theta + a^2)\,dr.$$

В частном случае нормального распределения $h(u) = (2\pi D)^{-1} e^{-u/2D}$ и формула для $f_3(r)$ принимает вид

$$f_3(r) = \frac{r}{2\pi D} e^{-\frac{r^2+a^2}{2D}} \int_0^{2\pi} e^{\frac{ar}{D}\cos\theta}\,d\theta.$$

Интеграл в этой формуле не сводится к элементарным функциям. Его можно выразить через бесселеву функцию

$$I_0(u) = \frac{1}{\pi}\int_0^{\pi} e^{u\cos\theta}\,d\theta = \frac{1}{2\pi}\int_0^{2\pi} e^{u\cos\theta}\,d\theta.$$

В результате получим

$$f_3(r) = \frac{r}{D}\, e^{-\frac{r^2+a^2}{2D}}\, I_0\left(\frac{ar}{D}\right). \qquad (40)$$

Распределение, определяемое этой формулой, называется *нецентральным распределением Рэлея.*

Пример 28. Найти плотность случайного вектора $Y = AX + c$, если известна плотность $f_1(x)$ вектора X.

В случае, когда векторы X и Y имеют одинаковую размерность n и ранг матрицы A равен n, уравнение $y = Ax + c$ имеет единственное решение $x = A^{-1}(y-c)$ и формула (31) дает

$$f_2(y) = \frac{1}{\sqrt{|A|^2}}\, f_1(A^{-1}(y-c)),$$

где $|A|$ — определитель матрицы A.

В частном случае нормального распределения вектора X (§ 4.4)

$$f_1(x) = \frac{1}{\sqrt{(2\pi)^n |K_x|}} \exp\left\{-\frac{1}{2}\left(x^T - m_x^T\right) K_x^{-1}(x - m_x)\right\}$$

и полученная формула дает

$$f_2(y) = \frac{1}{\sqrt{(2\pi)^n |K_x|\,|A|^2}} \times$$
$$\times \exp\left\{-\frac{1}{2}\left(y^T - c^T - m_x^T A^T\right) A^{-1T} K_x^{-1} A^{-1}(y - c - Am_x)\right\}.$$

Это нормальное распределение, причем математическое ожидание вектора Y равно $m_y = Am_x + c$, а его ковариационная матрица равна $K_y = AK_xA^T$ в полном соответствии с формулами (3.49) и (3.50).

Предоставляем читателю самостоятельно рассмотреть случай, когда размерность вектора Y и ранг матрицы A равны $m < n$, случай, когда размерность вектора Y равна $m \leqslant n$, а ранг матрицы A равен $r < m$, и случай, когда $m > n$. Доказать, что во всех этих случаях при нормальном распределении $f_1(x)$ распределение $f_2(y)$ тоже нормально (является вырожденным нормальным распределением при $r < m$ и при $m > n$).

Таким образом, *случайные величины, полученные в результате любых линейных преобразований нормально распределенных случайных величин, распределены нормально. Этот важный результат имеет большое значение для теории вероятностей и ее применений.* В частности, отсюда следует, что коэффициенты канонического разложения нормально распределенного случайного вектора, в частности его разложения по собственным векторам, распределены нормально (имеют совместное нормальное распределение). А так как из некоррелированности координат нормально распределенного слу-

чайного вектора следует их независимость, то коэффициенты канонического разложения нормально распределенного случайного вектора всегда независимы.

Из результатов этого примера вытекает также следующее обобщение теоремы § 4.4: *проекции нормально распределенного случайного вектора на любые подпространства распределены нормально.*

Пример 29. Случайный вектор X распределен нормально, причем ранг r его ковариационной матрицы K_x меньше его размерности n. Найти плотность вектора X.

Мы знаем (§ 3.3), что в этом случае вектор X имеет вырожденное нормальное распределение (4.62), которое полностью сосредоточено на r-мерном подпространстве, образованном первыми r собственными векторами матрицы K_x, сдвинутом на вектор m_x (если m_x не принадлежит этому r-мерному подпространству). На основании результатов предыдущего примера распределение вектора X можно выразить через распределение в этом подпространстве.

Пусть Φ — матрица, столбцы которой представляют собой, как и в п. 4.4.6, ортонормированные собственные векторы $\varphi_1, \ldots, \varphi_n$ матрицы K_x. Эта матрица ортогональна, вследствие чего ее определитель равен единице, $|\Phi| = 1$. На основании результатов предыдущего примера случайные величины $U_k = \varphi_k^T (X - m_x)$ $(k = 1, \ldots, r)$ независимы, имеют нулевые математические ожидания и дисперсии, равные соответственно r отличным от нуля собственным значениям $\lambda_1, \ldots, \lambda_r$ матрицы K_x. Поэтому распределение вектора $X^0 = X - m_x$ в r-мерном подпространстве, образованном собственными векторами $\varphi_1, \ldots, \varphi_r$, определяется плотностью

$$\frac{1}{\sqrt{(2\pi)^r \lambda_1 \ldots \lambda_r}} \exp\left\{-\frac{1}{2} \sum_{k=1}^{r} \frac{u_k^2}{\lambda_k}\right\}.$$

Случайные величины $U_k = \varphi_k^T X^0$ $(k = r+1, \ldots, n)$ имеют нулевые математические ожидания и нулевые дисперсии, вследствие чего они с вероятностью 1 равны 0. Поэтому их плотности представляют собой δ-функции $\delta(u_k)$ $(k = r+1, \ldots, n)$. Следовательно, плотность n-мерного случайного вектора $U = [U_1 \ldots U_n]^T$ определяется формулой

$$f_1(u) = \frac{1}{\sqrt{(2\pi)^r \lambda_1 \ldots \lambda_r}} \exp\left\{-\frac{1}{2} \sum_{k=1}^{r} \frac{u_k^2}{\lambda_k}\right\} \delta(u_{r+1}) \ldots \delta(u_n).$$

Это вырожденное нормальное распределение в пространстве значений случайного вектора X, в котором за оси координат приняты направления собственных векторов матрицы K_x. Пользуясь формулой (31), находим плотность случайного вектора X:

$$f_2(x) = \frac{1}{\sqrt{(2\pi)^r \lambda_1 \ldots \lambda_n}} \exp\left\{-\frac{1}{2} (x^T - m_x^T) \sum_{k=1}^{r} \frac{\varphi_k \varphi_k^T}{\lambda_k} (x - m_x)\right\} \times$$
$$\times\, \delta\left(\varphi_{r+1}^T (x - m_x)\right) \ldots \delta\left(\varphi_n^T (x - m_x)\right). \qquad (41)$$

Мы получили другую форму вырожденного нормального распределения, отличную от (4.62).

Предоставляем читателю самостоятельно показать, что в случае $r=n$ формула (41) совпадает с формулой (4.51), определяющей плотность n-мерного нормально распределенного вектора.

Пример 30. Пусть X и Y — две случайные величины (скалярные или векторные одной и той же размерности) с совместной плотностью $f_1(x, y)$. Найти плотность их суммы $Z=X+Y$.

Так как уравнение $z=x+y$ при любом z имеет единственное решение как относительно x, так и относительно y, то можно воспользоваться формулой (34). В результате получим

$$f_2(z) = \int_{-\infty}^{\infty} f_1(x, z-x)\,dx = \int_{-\infty}^{\infty} f_1(z-y, y)\,dy. \tag{42}$$

В частном случае независимых слагаемых $f_1(x, y)=g_1(x)\,g_2(y)$ и

$$f_2(z) = \int_{-\infty}^{\infty} g_1(x)\,g_2(z-x)\,dx = \int_{-\infty}^{\infty} g_1(z-y)\,g_2(y)\,dy. \tag{43}$$

Интегралы такого типа называются *сверткой* функций g_1 и g_2 и обозначаются g_1*g_2. Таким образом, *плотность суммы независимых случайных величин равна свертке плотностей слагаемых*, $f_2=g_1*g_2$.

Так как сумма неслучайных величин не может быть случайной величиной, а любая неслучайная величина имеет плотность, представляющую собой δ-функцию, то из (43) следует, что свертка двух δ-функций является δ-функцией:

$$\int_{-\infty}^{\infty} \delta(x-a)\,\delta(z-x-b)\,dx = \delta(z-a-b). \tag{44}$$

Это свойство δ-функции можно также вывести предельным переходом, с помощью которого δ-функция определена в приложении 1.

Из общего результата примера 28 следует, что свертка двух нормальных распределений представляет собой нормальное распределение. Предлагаем читателю самостоятельно получить этот результат по формуле (43).

Пример 31. Найти плотность суммы двух независимых случайных величин, имеющих χ^2-распределения с k и l степенями свободы соответственно (пример 10, формула (11)).

В данном случае

$$g_1(x) = \frac{1(x)}{2^{\frac{k}{2}}\,\Gamma\left(\frac{k}{2}\right)}\,x^{\frac{k}{2}-1}\,e^{-\frac{1}{2}x}, \quad g_2(y) = \frac{1(y)}{2^{\frac{l}{2}}\,\Gamma\left(\frac{l}{2}\right)}\,y^{\frac{l}{2}-1}\,e^{-\frac{1}{2}y}.$$

и формула (43) дает

$$f_2(z) = \frac{1(z)}{2^{\frac{k+l}{2}} \Gamma\left(\frac{k}{2}\right) \Gamma\left(\frac{l}{2}\right)} \int_0^z x^{\frac{k}{2}-1} (z-x)^{\frac{l}{2}-1} e^{-\frac{1}{2}x - \frac{1}{2}(z-x)} dx =$$

$$= \frac{1(z)}{2^{\frac{k+l}{2}} \Gamma\left(\frac{k}{2}\right) \Gamma\left(\frac{l}{2}\right)} z^{\frac{k+l}{2}-1} e^{-\frac{1}{2}z} \int_0^1 u^{\frac{k}{2}-1} (1-u)^{\frac{l}{2}-1} du.$$

Последний интеграл представляет собой так называемую *бета-функцию:*

$$\int_0^1 u^{p-1} (1-u)^{q-1} du = \mathrm{B}(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \qquad (45)$$

Пользуясь этой формулой, получаем

$$f_2(z) = \frac{1(z)}{2^{\frac{k+l}{2}} \Gamma\left(\frac{k+l}{2}\right)} z^{\frac{k+l}{2}-1} e^{-\frac{1}{2}z}.$$

Таким образом, *сумма независимых случайных величин, распределенных по закону* χ^2, *также имеет* χ^2*-распределение с числом степеней свободы, равным сумме чисел степеней свободы слагаемых.*

Пример 32. Найти распределение суммы Z независимых случайных величин X и Y, распределенных по закону Пуассона с параметрами λ и μ соответственно.

В данном случае

$$g_1(x) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta(x-k), \quad g_2(y) = \sum_{h=0}^{\infty} \frac{\mu^h}{h!} e^{-\mu} \delta(y-h)$$

и формула (43) дает

$$f_3(z) = \sum_{k, h=0}^{\infty} \frac{\lambda^k \mu^h}{k!\,h!} e^{-\lambda-\mu} \int_{-\infty}^{\infty} \delta(x-k)\,\delta(z-x-h)\,dx =$$

$$= \sum_{k, h=0}^{\infty} \frac{\lambda^k \mu^h}{k!\,h!} e^{-\lambda-\mu} \delta(z-k-h).$$

Собрав вместе слагаемые, соответствующие одинаковым значениям суммы $k+h=m$, имея в виду, что

$$\sum_{k+h=m} \frac{\lambda^k \mu^h}{k!\,h!} = \frac{1}{m!} \sum_{k=0}^{m} \frac{m!}{k!\,(m-k)!} \lambda^k \mu^{m-k} = \frac{(\lambda+\mu)^m}{m!},$$

получим

$$f_3(z) = \sum_{m=0}^{\infty} \frac{(\lambda+\mu)^m}{m!} e^{-\lambda-\mu} \delta(z-m).$$

Таким образом, *сумма независимых случайных величин, распределенных по закону Пуассона, также распределена по закону Пуассона с параметром, равным сумме параметров распределений слагаемых.*

Пример 33. Вектор X имеет нормальное распределение с шаровой симметрией в n-мерном пространстве, причем дисперсии всех координат вектора X равны 1. Найти распределение случайной величины $Y = X^T A X$, где A — неотрицательно определенная симметричная матрица ранга $r < n$, обладающая свойством идемпотентности $A^2 = A$.

Матрица A представляет собой матрицу ортогонального проектирования на некоторое r-мерное подпространство. А так как проекция нормально распределенного вектора на любое подпространство распределена нормально, причем шаровая симметрия распределения при ортогональном проектировании сохраняется, то на основании результатов примера 10 интуитивно ясно, что распределение величины Y представляет собой χ^2-распределение с r степенями свободы. Чтобы строго доказать это, приведем матрицу A ортогональным преобразованием к диагональной форме. Так как ранг матрицы A равен $r < n$, то $n-r$ собственных значений матрицы A будут равны нулю, $\lambda_{r+1} = \ldots = \lambda_n = 0$. Пусть $\varphi_1, \ldots, \varphi_n$ — ортонормированные собственные векторы матрицы A, Φ — $n \times r$-матрица, столбцами которой служат собственные векторы $\varphi_1, \ldots, \varphi_r$, соответствующие отличным от нуля собственным значениям $\lambda_1, \ldots, \lambda_r$, Φ_0 — $n \times (n-r)$-матрица, столбцами которой служат собственные векторы $\varphi_{r+1}, \ldots, \varphi_n$, соответствующие нулевому собственному значению, Λ — диагональная матрица порядка r с элементами $\lambda_1, \ldots, \lambda_r$. Тогда можем написать

$$A\Phi = \Phi\Lambda, \quad A\Phi_0 = 0.$$

Так как по условию $A^2 = A$, то на основании первого из этих соотношений $A^2\Phi = \Phi\Lambda$, а с другой стороны, $A^2\Phi = A(A\Phi) = A\Phi\Lambda = \Phi\Lambda^2$. Таким образом, $\Lambda = \Lambda^2$, что для диагональной матрицы возможно, только если Λ представляет собой единичную матрицу порядка r, $\Lambda = I_r$. Таким образом, $A\Phi = \Phi$ и матрица A выражается формулой

$$A = [\Phi \ \ \Phi_0] \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Phi^T \\ \Phi_0^T \end{bmatrix} = \Phi\Phi^T.$$

Рассмотрим теперь случайный вектор $U = \Phi^T X$. Его ковариационная матрица на основании (3.50) равна $K_u = \Phi^T K_x \Phi = \Phi^T\Phi$, так как по условиям задачи $K_x = I$. Но $\Phi^T\Phi = I_r$ в силу ортонормальности собственных векторов $\varphi_1, \ldots, \varphi_n$. Поэтому $K_u = I_r$. Наконец,

$$U^T K_u^{-1} U = U^T U = X^T \Phi\Phi^T X = X^T A X = Y.$$

Таким образом, интересующая нас случайная величина Y представляет собой квадратичную форму, которая с коэффициентом 1/2 стоит в показателе степени в выражении нормальной плотности r-мерного вектора U. Следовательно, по доказанному в примере 10 величина Y

имеет χ^2-распределение с r степенями свободы:

$$f_2(y) = \frac{1(y)}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)} y^{\frac{r}{2}-1} e^{-\frac{1}{2}y}.$$

Доказать, что и в более общем случае, когда X имеет распределение $N(0, K_x)$ (может быть, и вырожденное), случайная величина $Y = X^T AX$ имеет χ^2-распределение с r степенями свободы, если матрица $B = K_x^{1/2} A K_x^{1/2}$ представляет собой матрицу ортогонального проектирования на некоторое r-мерное подпространство (т. е. имеет ранг r и обладает свойством идемпотентности $B^2 = B$).

Пример 34. В условиях предыдущего примера найти распределение случайной величины $Y = X^T X - n\overline{X}^2$, $\overline{X} = (1/n) \sum\limits_{p=1}^{n} X_p$.

В данном случае

$$Y = \sum_{p=1}^{n} X_p^2 - \frac{1}{n} \sum_{p,\, q=1}^{n} X_p X_q = X^T AX,$$

где

$$A = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} \dots & -\frac{1}{n} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ -\frac{1}{n} & -\frac{1}{n} \dots & 1-\frac{1}{n} \end{bmatrix}.$$

Эта матрица имеет ранг $r = n-1$, и $A^2 = A$. Следовательно, величина Y, согласно результатам примера 33, имеет χ^2-распределение с $r = n-1$ степенями свободы:

$$f_2(y) = \frac{1(y)}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} y^{\frac{n-3}{2}} e^{-\frac{1}{2}y}.$$

Пример 35. Пользуясь формулами (31), (34) — (38), показать, что формула (3.6) дает для математического ожидания функции $Y = \varphi(X)$ случайной величины X то же значение, что и формула (3.3):

$$MY = M\varphi(X) = \int_{-\infty}^{\infty} \varphi(x) f_1(x)\, dx = \int_{-\infty}^{\infty} y f_2(y)\, dy.$$

Пример 36. С помощью формул этой главы доказать теорему п. 4.2.5: *если величины $X_1, \dots, X_n$ независимы, то и случайные величины $Y_1 = \varphi_1(X_1), \dots, Y_n = \varphi_n(X_n)$ независимы при любых функциях* $\varphi_1, \dots, \varphi_n$ (с помощью формул (31), (34) — (38) это доказывается для дифференцируемых функций $\varphi_1, \dots, \varphi_n$, а с помощью формулы (9) — для любых измеримых функций $\varphi_1, \dots, \varphi_n$).

Пример 37. Если расположить значения n независимых скалярных случайных величин $X_1, \ldots, X_n$, имеющих одну и ту же функцию распределения $F(x)$, в порядке возрастания, то ν-е по порядку значение называется *ν-м нижним значением*, а ν-е в порядке убывания называется *ν-м верхним значением*. Найти распределения ν-го нижнего U и ν-го верхнего V значений.

В данном случае проще всего применить метод сравнения элементов вероятности. Вычислим элемент вероятности величины U, соответствующий бесконечно малому интервалу $(u, u+du)$ (т. е. вероятность попадания ν-го нижнего значения U в этот интервал). Чтобы величина U попала в интервал $[u, u+du)$, необходимо и достаточно, чтобы какие-нибудь $\nu-1$ из n величин $X_1, \ldots, X_n$ приняли значения, меньшие чем u, одно значение попало в интервал $[u, u+du)$, а остальные $n-\nu$ приняли значения, не меньшие чем $u+du$.

Таким образом, мы имеем схему повторения опытов с тремя несовместными событиями, образующими полную группу, $A_1=\{X<u\}$, $A_2=\{X \geqslant u+du\}$ и $A_3=\{u \leqslant X < u+du\}$, вероятности которых при одном опыте равны соответственно $F(u)$, $1-F(u)$ и $f(u)\,du$ (с точностью до бесконечно малых высших порядков). Поэтому для вычисления искомой вероятности можно применить формулу (1.36) для полиномиального распределения при $r=3$. Тогда получим

$$f_1(u)\,du=\frac{n!}{(\nu-1)!\,(n-\nu)!}F^{\nu-1}(u)\,[1-F(u)]^{n-\nu}f(u)\,du.$$

Эта формула определяет плотность ν-го нижнего значения U:

$$f_1(u)=\frac{n!}{(\nu-1)!\,(n-\nu)!}F^{\nu-1}(u)\,[1-F(u)]^{n-\nu}f(u). \tag{46}$$

Точно так же находим плотность ν-го верхнего значения V:

$$f_2(v)=\frac{n!}{(\nu-1)!\,(n-\nu)!}F^{n-\nu}(v)\,[1-F(v)]^{\nu-1}f(v). \tag{47}$$

В частности, при $\nu=1$ отсюда получаются формулы для плотностей наименьшего и наибольшего значений $U=\min\limits_{1\leqslant k\leqslant n} X_k$, $V=\max\limits_{1\leqslant k\leqslant n} X_k$:

$$f_1(u)=n\,[1-F(u)]^{n-1}f(u), \quad f_2(v)=nF^{n-1}(v)\,f(v). \tag{48}$$

Для нахождения совместной плотности ν-го нижнего и ν-го верхнего значений U, V при $\nu \leqslant n/2$ достаточно заметить, что для попадания U в интервал $(u, u+du)$, а V в интервал $(v, v+dv)$, $v>u$, необходимо и достаточно, чтобы какие-нибудь $\nu-1$ из величин $X_1, \ldots, X_n$ приняли значения, меньшие u, какие-нибудь $\nu-1$ других приняли значения, не меньшие $v+dv$, $n-2\nu$ попали в интервал $[u+du, v)$, одно попало в интервал $[u, u+du)$ и одно в интервал $[v, v+dv)$. Таким образом, задача сводится к схеме повторения опытов с пятью событиями, вероятности которых равны $F(u)$, $1-F(v)$, $F(v)-F(u)$, $f(u)\,du$, $f(v)\,dv$. Пользуясь формулой (1.36) при $r=5$, получаем после сокращения на $du\,dv$ следующее выражение для совместной плотности ν-го нижнего и ν-го верхнего значений U, V:

$$f(u, v)=\frac{n!}{[(\nu-1)!]^2\,(n-2\nu)!}F^{\nu-1}(u)\,[F(v)-$$
$$-F(u)]^{n-2\nu}\,[1-F(v)]^{\nu-1}f(u)\,f(v)\,1(v-u). \tag{49}$$

Последний множитель $1(v-u)$ добавлен потому, что событие $V < U$ при $\nu \leqslant n/2$ невозможно, вследствие чего $f(u, v) = 0$ при $v < u$.

В частном случае при $\nu = 1$ из (49) получается формула для совместной плотности наименьшего и наибольшего значений $U = \min\limits_{1 \leqslant k \leqslant n} X_k$, $V = \max\limits_{1 \leqslant k \leqslant n} X_k$:

$$f(u, v) = n(n-1)[F(v) - F(u)]^{n-2} f(u) f(v) 1(v-u). \qquad (50)$$

Пример 38. В условиях предыдущего примера найти распределение разности между ν-м верхним и ν-м нижним значениями $S = V - U$.

Так как уравнение $s = v - u$ при любом s имеет единственное решение $v = u + s$, то на основании (34) плотность величины S определяется формулой

$$f_3(s) = \int_{-\infty}^{\infty} f(u, u+s)\, du. \qquad (51)$$

Подставив сюда выражение (49) функции $f(u, v)$, получим

$$f_3(s) = \frac{n!\, 1(s)}{[(\nu-1)!]^2 (n-2\nu)!} \int_{-\infty}^{\infty} F^{\nu-1}(u) [F(u+s) - F(u)]^{n-2\nu} [1 - F(u+s)]^{\nu-1} f(u) f(u+s)\, du. \qquad (52)$$

В частном случае при $\nu = 1$ отсюда получаем плотность так называемой *широты разброса* n независимых случайных величин $S = \max\limits_{1 \leqslant k \leqslant n} X_k - \min\limits_{1 \leqslant k \leqslant n} X_k$ (*широты выборки* в задачах статистики):

$$f_3(s) = n(n-1) 1(s) \int_{-\infty}^{\infty} [F(u+s) - F(u)] f(u) f(u+s)\, ds. \qquad (53)$$

3.3. Применение дельта-функции. Еще один метод нахождения распределений функций случайных величин основан на применении δ-функции.

▶ Если $Y = \varphi(X)$, то величина Y при любом значении x величины X имеет единственное возможное значение $\varphi(x)$, и условная вероятность этого значения при данном x равна 1. Поэтому условная плотность величины Y при $X = x$ представляет собой δ-функцию:

$$f_2(y \mid x) = \delta(y - \varphi(x)).$$

Следовательно, совместная плотность величин X и $Y = \varphi(X)$ определяется формулой

$$f(x, y) = f_1(x)\, \delta(y - \varphi(x)).$$

Отсюда, выполнив интегрирование по x в соответствии с (4.9), находим плотность случайной величины Y:

$$f_2(y)=\int_{-\infty}^{\infty} f_1(x)\,\delta(y-\varphi(x))\,dx. \blacktriangleleft \qquad (54)$$

Эта формула определяет плотность величины $Y=\varphi(X)$ как в случае скалярных, так и в случае векторных величин X и Y. Она особенно удобна для нахождения плотности функции случайной величины в случаях, когда функция $\varphi(x)$ имеет постоянные значения в некоторых областях пространства значений величины X, вероятности попадания в которые отличны от нуля. В таких случаях формула (54) автоматически дает линейную комбинацию соответствующих δ-функций в выражении плотности $f_2(y)$, в то время как формулы (31), (34)—(38), так же как и метод сравнения элементов вероятности, оказываются неприменимыми (следует отметить, что формула (9) справедлива и в таких случаях).

Предоставляем читателю самостоятельно вывести из (54) формулы (31), (34)—(38), выполнив интегрирование после соответствующей замены переменных и, если потребуется, разбиения области интегрирования на части.

Пример 39. Найти распределение выходного сигнала ограничителя (пример 2.15), если дана плотность $f_1(x)$ входного сигнала X.

В примере 2.15 мы нашли плотность выходного сигнала Y, определив сначала его функцию распределения (т. е. применением формулы (9) к данной частной задаче). Сейчас мы решим эту задачу, пользуясь формулой (54):

$$f_2(y)=$$
$$=\int_{-\infty}^{a} f_1(x)\,\delta(y+a)\,dx+\int_{-a}^{a} f_1(x)\,\delta(y-x)\,dx+\int_{a}^{\infty} f_1(x)\,\delta(y-a)\,dx=$$
$$=f_1(y)+\delta(y+a)\int_{-\infty}^{a} f_1(x)\,dx+\delta(y-a)\int_{a}^{\infty} f_1(x)\,dx.$$

Пример 40. Найти распределение суммы Z независимых случайных величин X и Y, распределенных равномерно на интервалах $(-a, a)$ и $(-b, b)$ соответственно, $b>a$.

По формуле (54) находим

$$f_2(z)=\frac{1}{4ab}\int_{-a}^{a} dx\int_{-b}^{b}\delta(z-x-y)\,dy.$$

Так как точка $z-x$ расположена в интервале $(-b, b)$ только при $z-x+b>0$, $z-x-b<0$, то интеграл по переменной y равен разности единичных ступенчатых функций $1(z-x+b)-1(z-x-b)$ и

$$f_2(z)=\frac{1}{4ab}\int_{-a}^{a}[1(z-x+b)-1(z-x-b)]\,dx.$$

Подынтегральная функция здесь отлична от нуля только при $z-b< x<z+b$. Интервал $(z-b, z+b)$ не пересекается с интервалом интегрирования $(-a, a)$ при $z<-a-b$ и при $z>a+b$,

$$\begin{aligned}
&(z-b,\ z+b)\cap(-a,\ a)=(-a,\ z+b) && \text{при} \quad z\in(-a-b,\ a-b),\\
&(z-b,\ z+b)\cap(-a,\ a)=(-a,\ a) && \text{при} \quad z\in(a-b,\ b-a),\\
&(z-b,\ z+b)\cap(-a,\ a)=(z-b,\ a) && \text{при} \quad z\in(b-a,\ a+b).
\end{aligned}$$

Поэтому $f_2(z)=0$ при $|z|>a+b$,

$$f_2(z)=\begin{cases}\dfrac{1}{4ab}(a+b+z) & \text{при} \quad z\in(-a-b,\ a-b),\\[2mm] \dfrac{1}{2b} & \text{при} \quad z\in(a-b,\ b-a),\\[2mm] \dfrac{1}{4ab}(a+b-z) & \text{при} \quad z\in(b-a,\ a+b).\end{cases}$$

График этой плотности показан на рис. 25.

3.4. Применение характеристических функций. В некоторых случаях для нахождения распределений функций случайных величин целесообразно применение метода характеристических функций.

Рассмотрим случайную величину $Y=\varphi(X)$, где $\varphi(x)$ — любая функция (измеримая, см. п. 2.2.). Согласно определению (4.63) характеристическая функция $g_2(\lambda)$ величины Y дается формулой

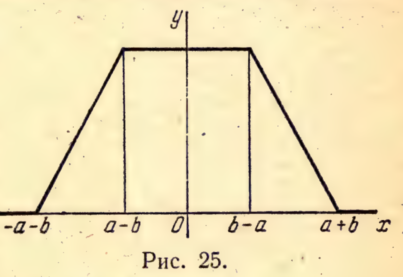


Рис. 25.

$$g_2(\lambda)=Me^{i\lambda^T Y}=Me^{i\lambda^T\varphi(X)}=\int_{-\infty}^{\infty}e^{i\lambda^T\varphi(x)}f_1(x)\,dx, \quad (55)$$

где $f_1(x)$ — плотность случайной величины X. Определив таким путем характеристическую функцию величины Y, можно по формуле (4.72) найти ее плотность:

$$f_2(y)=\frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}e^{-i\lambda^T y}g_2(\lambda)\,d\lambda. \quad (56)$$

Подставив выражение (55) в (56), изменив формально порядок интегрирования и приняв во внимание формулу (15) приложения 1, вновь получим (54).

Пример 41. Найти распределение суммы двух независимых величин X и Y, распределенных равномерно в интервалах $(-a, a)$ и $(-b, b)$ соответственно, $a < b$.

Формулы (55) и (56) дают

$$f_3(z) = \frac{1}{8\pi ab}\int_{-\infty}^{\infty} e^{-i\lambda z}\,d\lambda \int_{-a}^{a}\int_{-b}^{b} e^{i\lambda(x+y)}\,dx\,dy =$$

$$= \frac{1}{2\pi ab}\int_{-\infty}^{\infty} e^{-i\lambda z}\frac{\sin\lambda a\sin\lambda b}{\lambda^2}\,d\lambda = \frac{1}{\pi ab}\int_{-\infty}^{\infty}\frac{\sin\lambda a\sin\lambda b\cos\lambda z}{\lambda^2}\,d\lambda =$$

$$= \frac{1}{4\pi ab}\int_{0}^{\infty}[\cos(a-b+z)\lambda + \cos(a-b-z)\lambda -$$

$$-\cos(a+b+z)\lambda - \cos(a+b-z)\lambda]\frac{d\lambda}{\lambda^2}.$$

Отсюда на основании известной формулы

$$\int_0^{\infty}\frac{1-\cos\alpha u}{u^2}\,du = \frac{\pi}{2}|\alpha|$$

получаем

$$f_3(z) = \frac{1}{8ab}(|a+b+z| + |a+b-z| - |a-b+z|\,a-b-z|).$$

Этот результат совпадает с результатом примера 40.

Пример 42. Координаты случайного вектора X независимы и имеют распределение $N(0, D)$. Найти совместное распределение величин $\overline{X}$ и $S = X^T X - n\overline{X}^2$, где $\overline{X}$ — среднее арифметическое координат вектора X.

По формуле (55) находим

$$g(\lambda_1, \lambda_2) = Me^{i\lambda_1\overline{X} + i\lambda_2 S} =$$

$$= \frac{1}{(2\pi D)^{n/2}}\int_{-\infty}^{\infty}\exp\left\{i\lambda_1\overline{x} + i\lambda_2(|x|^2 - n\overline{x}^2) - \frac{1}{2D}|x|^2\right\}dx.$$

Замена переменных $y = Ax$, где A — ортогональная матрица, все элементы первой строки которой равны $1/\sqrt{n}$, дает $\sqrt{n}\,\overline{x} = y_1$, и интеграл преобразуется к произведению n интегралов

$$g(\lambda_1, \lambda_2) =$$

$$= \frac{1}{(2\pi D)^{n/2}}\int_{-\infty}^{\infty}\exp\left\{\frac{i\lambda_1 y}{\sqrt{n}} - \frac{y^2}{2D}\right\}dy\left[\int_{-\infty}^{\infty}\exp\left\{\frac{1-2iD\lambda_2}{2D}y^2\right\}dy\right]^{n-1}.$$

Пользуясь формулой (1) приложения 2, получим

$$g(\lambda_1, \lambda_2) = e^{-\frac{D}{2n}\lambda_1^2}(1-2iD\lambda_2)^{-\frac{n-1}{2}}.$$

Отсюда в силу свойств 6) и 3) характеристических функций и формул (4.68) и (4.69) при $\mu = n/2-1$ и $k=1/2$ следует, что величины $\bar{X}$ и S независимы, $\bar{X}$ имеет распределение $N(0, D/n)$, а S/D — χ^2-распределение с $k=n-1$ степенями свободы.

Пример 43. Строки $n\times m$-матрицы X представляют собой независимые m-мерные векторы, имеющие одно и то же распределение $N(0, K)$, $\bar{X}$ — матрица-строка, элементами которой служат средние арифметические элементов соответствующих столбцов матрицы X (пример 15). Найти совместное распределение величин $\bar{X}$ и $S = X^T X - n\bar{X}^T\bar{X}$.

По формуле (55) находим

$$g(\lambda_1, \lambda_2) = Me^{i\bar{X}\lambda_1 + i\,\mathrm{tr}\,\lambda_2 S} =$$

$$= \frac{1}{[(2\pi)^m |K|]^{n/2}} \int_{-\infty}^{\infty} \exp\left\{i\bar{x}\lambda_1 + \mathrm{tr}\,\lambda_2(x^T x - n\bar{x}^T\bar{x}) - \frac{1}{2}\,\mathrm{tr}\,xK^{-1}x^T\right\} dx,$$

где λ_1 — матрица-столбец $m\times 1$, а λ_2 — симметричная $m\times m$-матрица. Применив ортогональное преобразование предыдущего примера к каждому столбцу матрицы x, получим

$$g(\lambda_1, \lambda_2) = \frac{1}{[(2\pi)^m |K|]^{n/2}} \int_{-\infty}^{\infty} \exp\left\{\frac{iy\lambda_1}{\sqrt{n}} - \frac{1}{2}\,yK^{-1}y^T\right\} dy \times$$

$$\times \left[\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\,y(K^{-1}-2i\lambda_2)\,y^T\right\} dy\right]^{n-1},$$

где y — матрица-строка $1\times m$. Пользуясь формулой (2) приложения 2, получим окончательно

$$g(\lambda_1, \lambda_2) = e^{-\lambda_1^T K\lambda_1/2n}\,[|K|\,|K^{-1}-2i\lambda_2|]^{-(n-1)/2} =$$

$$= e^{-\lambda_1^T K\lambda_1/2n}\,|I-2iK\lambda_2|^{-(n-1)/2}.$$

Отсюда в силу свойств 6) и 3) характеристических функций и (4.70) следует, что величины $\bar{X}$ и S независимы, причем $\bar{X}$ имеет распределение $N(0, K/n)$, а характеристическая функция величины S определяется формулой

$$g_S(\lambda) = |I - 2iK\lambda|^{-(n-1)/2}, \tag{57}$$

где λ — симметричная $m\times m$-матрица. Чтобы найти распределение величины S, остается заметить, что $g_S(\lambda)$ абсолютно интегрируема, и применить формулу (56). В результате после довольно громоздких выкладок получим для S распределение Уишарта, определяемое плотностью $w_{nm}(s)$ (пример 18) [105].

Пример 44. Совершенно так же в условиях примера 43 находим распределение величины $S' = X^T X$:

$$g_{s'}(\lambda) = \frac{1}{[(2\pi)^m |K|]^{n/2}} \int_{-\infty}^{\infty} e^{i \operatorname{tr} \lambda x^T x - \frac{1}{2} \operatorname{tr} x K^{-1} x^T} dx =$$

$$= |I - 2iK\lambda|^{-\frac{n}{2}}.$$

Сравнивая эту формулу с (57), видим, что величина S' имеет распределение Уишарта $w_{n+1, m}(s')$.

3.5. Метод моментов. Если функция $\varphi(x)$ ограничена, то для нахождения распределения величины $Y = \varphi(X)$ по данному распределению величины X можно применить метод моментов.

По доказанному в п. 4.5.3 распределение ограниченной случайной величины полностью определяется ее моментами. Поэтому в данном случае достаточно вычислить моменты величины $Y = \varphi(X)$.

Пример 45. В условиях примеров 43 и 44 найти распределение отношения определителей матриц S и S', $V = |S|/|S'|$.

Заметив, что

$$|S'| = |S| (1 + n\bar{X} S^{-1} \bar{X}^T), \tag{58}$$

и приняв во внимание положительную определенность матрицы S, приходим к выводу, что $|S'| \geqslant |S| \geqslant 0$. Таким образом, распределение величины V полностью сосредоточено на интервале $[0, 1]$. Следовательно, распределение величины V полностью определяется ее моментами.

Для вычисления моментов величины V выведем сначала формулы для моментов величин $|S|$ и $|S'|$. На основании формулы (21), определяющей плотность $w_{nm}(s)$ матрицы S,

$$M |S|^r = \frac{c_{nm}}{(2^m |K|)^{\frac{n-1}{2}}} \int |s|^{r + \frac{n-m}{2} - 1} e^{-\frac{1}{2} \operatorname{tr} K^{-1} s} ds.$$

Заметив, что подынтегральная функция, умноженная на $c_{n+2r, m} (2^m |K|)^{-(n+2r-1)/2}$, представляет собой плотность $w_{n+2r, m}(s)$, получаем

$$M |S|^r = (2^m |K|)^r \frac{c_{nm}}{c_{n+2r, m}} = (2^m |K|)^r \prod_{i=1}^{m} \frac{\Gamma\left(\frac{n+2r-i}{2}\right)}{\Gamma\left(\frac{n-i}{2}\right)} \tag{59}$$

при любом $r > -(n-m)/2$. Так как распределение матрицы S' по доказанному в примере 44 определяется плотностью $w_{n+1, m}(s')$, то для нахождения моментов определителя $|S'|$ достаточно заменить

в (59) n на $n+1$. В результате будем иметь

$$M|S'|^r=(2^m|K|)^r\frac{c_{n+1\ m}}{c_{n+2r+1,\ m}}=(2^m|K|)^r\prod_{i=1}^{m}\frac{\Gamma\left(\frac{n+2r+1-i}{2}\right)}{\Gamma\left(\frac{n+1-i}{2}\right)} \qquad (60)$$

при любом $r>-(n+1-m)/2$.

Перейдем теперь к вычислению моментов величины $V=|S|/|S'|$. Заметив, что в силу (58) $V=(1+n\overline{X}S^{-1}\overline{X}^T)^{-1}$, т. е. представляет собой функцию величин $\overline{X}$ и S, пользуясь найденной в примере 43 совместной плотностью $\overline{X}$ и S, можем написать

$$MV^r=\frac{c_{nm}}{(2^m|K|)^{n/2}}\left(\frac{n}{\pi}\right)^{\frac{m}{2}}\iint\frac{|s|^r}{|s'|^r}|s|^{\frac{n-m}{2}-1}\times$$
$$\times e^{-\frac{1}{2}\operatorname{tr}K^{-1}s-\frac{n}{2}\bar{x}K^{-1}\bar{x}^T}d\bar{x}\,ds=$$
$$=\frac{c_{nm}}{(2^m|K|)^{n/2}\sqrt{\pi^m}}\iint\frac{1}{|s'|^r}|s|^{\frac{n+2r-m}{2}-1}\times$$
$$\times e^{-\frac{1}{2}\operatorname{tr}K^{-1}s-\frac{1}{2}zK^{-1}z^T}dz\,ds.$$

Последний интеграл, умноженный на

$$c_{n+2r,\ m}(2^m|K|)^{-\frac{n+2r}{2}}\pi^{-\frac{m}{2}},$$

представляет собой $M|S'|^{-r}$ в случае, когда n заменено на $n+2r$. Следовательно, для вычисления этой величины достаточно применить формулу (60), заменив в ней n на $n+2r$ и r на $-r$. В результате получим

$$MV^r=\frac{c_{nm}}{c_{n+2r,\ m}}\frac{c_{n+2r+1,\ m}}{c_{n+1,\ m}}=\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-m}{2}+r\right)}{\Gamma\left(\frac{n-m}{2}\right)\Gamma\left(\frac{n}{2}+r\right)}.$$

Сравнив эту формулу с формулой примера 3.19 для моментов β-распределения, приходим к выводу, что величина V имеет β-распределение с параметрами $p=(n-m)/2$ и $q=m/2$. Этот результат совпадает с результатом примера 18. Однако там этот результат был получен для произвольной функции $p(u)$ примера 15. Здесь же он получен для частного случая нормального распределения матрицы X.

Заметим в заключение, что в силу (58) величина V связана с обобщенным стьюдентовым отношением T (пример 15) зависимостью

$$V=\left(1+\frac{T^2}{n-1}\right)^{-1}.$$

Пользуясь этой зависимостью и найденным распределением величины V, можно легко получить результат примера 15 для частного случая нормального распределения матрицы X.

§ 4. Предельные теоремы

4.1. Простейшая предельная теорема. Метод характеристических функций удобен для нахождения предельных распределений функций неограниченно возрастающего числа случайных величин. Мы применим его для доказательства простейшей предельной теоремы.

Если $\{X_k\}$ — последовательность независимых величин, имеющих одно и то же распределение с математическим ожиданием m и ковариационной матрицей K, то последовательность распределений случайных величин

$$Z_n = \frac{1}{\sqrt{n}}(X_1^0 + \ldots + X_n^0), \quad X_k^0 = X_k - m \quad (n = 1, 2, \ldots),$$

сходится к нормальному распределению $N(0, K)$.

▶ Пусть $g(\lambda)$ — характеристическая функция каждой из величин X_k. На основании (4.65) и (4.66) характеристическая функция $g_n(\lambda)$ величины Z_n определяется формулой

$$g_n(\lambda) = \left[\exp\left\{-\frac{i\lambda^T m}{\sqrt{n}}\right\} g\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n.$$

Отсюда находим

$$\ln g_n(\lambda) = n\left[-\frac{i\lambda^T m}{\sqrt{n}} + \ln g\left(\frac{\lambda}{\sqrt{n}}\right)\right]. \qquad (61)$$

Но на основании (4.92)

$$\ln g\left(\frac{\lambda}{\sqrt{n}}\right) = \frac{i\lambda^T m}{\sqrt{n}} - \frac{\lambda^T K \lambda}{2n} + o\left(\frac{|\lambda|^2}{n}\right).$$

Подставив это выражение в (61), получим

$$\ln g_n(\lambda) = -\frac{1}{2}\lambda^T K \lambda + n o\left(\frac{|\lambda|^2}{n}\right).$$

Отсюда видно, что

$$\ln g_n(\lambda) \to -\frac{1}{2}\lambda^T K \lambda \quad \text{при} \quad n \to \infty.$$

Таким образом, $g_n(\lambda)$ стремится к характеристической функции нормального распределения $N(0, K)$ при каждом λ. Можно доказать, что отсюда следует сходимость последовательности распределений величин Z_n к $N(0, K)$ [17, 44]. ◀

Доказанная теорема, очевидно, справедлива и в том случае, когда ранг ковариационной матрицы K меньше размерности m величин X_ν. В этом случае предельное распределение $N(0, K)$ будет вырожденным.

Пример 46. В примере 3.11 было показано, что биномиальное распределение можно рассматривать как распределение суммы независимых величин $X_1, \ldots, X_n$, каждая из которых имеет два возможных значения 0 и 1 с вероятностями $q = 1 - p$ и p соответственно. Каждая из этих величин имеет математическое ожидание $m = p$ и дисперсию $D = pq$. На основании доказанной теоремы распределение случайной величины

$$Z_n = \frac{1}{\sqrt{n}} (X_1 + \ldots + X_n - np)$$

стремится к $N(0, pq)$ при $n \to \infty$ (теорема Лапласа).

Пример 47. Так как среднее арифметическое $\bar{X}$ независимых одинаково распределенных случайных величин $X_1, \ldots, X_n$ связано с величиной Z_n соотношением $\bar{X} = Z_n/\sqrt{n} + m$, то по доказанной теореме величина $\bar{X}$ распределена асимптотически нормально $N(m, K/n)$.

Пример 48. В примере 3.13 было показано, что случайный вектор $Z^{(i)}$, координатами которого служат числа появлений значений $x_1, \ldots, x_N$ дискретной случайной величины X в одном i-м опыте, имеет математическое ожидание $p = [p_1 \ldots p_N]^T$ и ковариационную матрицу $K = \Lambda - pp^T$, где $p_1, \ldots, p_N$ — вероятности значений $x_1, \ldots$ $\ldots, x_N$ соответственно, а Λ — диагональная матрица с элементами $p_1, \ldots, p_N$. По доказанной теореме случайный вектор

$$U = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z^{(i)} - p) = \frac{Z - np}{\sqrt{n}},$$

где Z — вектор, координатами которого служат числа появлений $Z_1, \ldots, Z_N$ значений $x_1, \ldots, x_N$ при n опытах, распределен асимптотически нормально $N(0, \Lambda - pp^T)$. Отсюда следует, что вектор частот $\hat{P} = [\hat{P}_1 \ldots \hat{P}_N]^T$ значений $x_1, \ldots, x_N$, равный $\hat{P} = U/\sqrt{n} + p$, распределен асимптотически нормально $N(p, (\Lambda - pp^T)/n)$.

Рассмотрим теперь случайный вектор V с координатами $V_\nu = (\hat{P}_\nu - p_\nu)\sqrt{n/p_\nu}$ $(\nu = 1, \ldots, N)$. Математическое ожидание вектора V равно нулю, а его ковариационная матрица K_v определяется формулой

$$K_v = I - qq^T, \qquad q = [\sqrt{p_1} \ldots \sqrt{p_N}]^T.$$

Согласно полученному результату распределение вектора V стремится к $N(0, K_v)$ при $n \to \infty$. Найдем асимптотическое распределение случайной величины

$$Y = n \sum_{\nu=1}^{N} \frac{(\hat{P}_\nu - p_\nu)^2}{p_\nu} = V^T V$$

при $n \to \infty$. Для этого воспользуемся последним результатом примера 33. В данном случае матрица A представляет собой единичную матрицу порядка N, а $B = K_v^{1/2} I K_v^{1/2} = K_v = I - qq^T$ имеет ранг $N - 1$ и представляет собой матрицу ортогонального проектирования, так как $q^T q = \sum (\sqrt{p_\nu})^2 = \sum p_\nu = 1$ и, следовательно, $(qq^T)^2 = q(q^T q) q^T = qq^T$ и

$$K_v^2 = I - 2qq^T + (qq^T)^2 = I - qq^T = K_v.$$

Согласно последнему результату примера 33 величина $Y = V^T V$ имеет асимптотическое распределение χ^2 с $N-1$ степенями свободы.

Координаты вектора V связаны очевидным скалярным соотношением $q^T V = 0$, вытекающим из равенств $\sum \hat{P}_\nu = \sum p_\nu = 1$ (напомним, что q — вектор с координатами $\sqrt{p_1}, \ldots, \sqrt{p_N}$). Именно поэтому ковариационная матрица вектора V имеет ранг $N-1$.

Теорема о сходимости распределения величины $V^T V$ к χ^2-распределению с $N-1$ степенями свободы была доказана К. Пирсоном [72, 73].

4.2. Значение предельных теорем. Из доказанной предельной теоремы следует, что распределение суммы случайных величин $X_1, \ldots, X_n$ при достаточно большом n как угодно близко к нормальному $N(\sum m_\nu, nK)$.

Теорема о сходимости последовательности распределений надлежащим образом нормированных сумм центрированных случайных величин к нормальному распределению при весьма общих условиях распространяется на случай независимых величин X_ν с различными распределениями, а также на случай зависимых величин X_ν [54, 55, 17, 44, 120].

Предельные теоремы образуют обширный раздел современной теории вероятностей. Они объясняют причины широкого распространения нормального распределения и механизм его формирования. На основании предельных теорем можно утверждать, что во всех случаях, когда случайная величина образуется в результате суммирования большого числа независимых или слабо зависимых случайных величин, дисперсия каждой из которых мала по сравнению с дисперсией суммы, распределение этой величины оказывается практически нормальным.

Глава 6

ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

§ 1. Основные задачи математической статистики

1.1. Определение статистических характеристик из опыта. Для получения исходных данных для вероятностных расчетов, так же как и для получения исходных данных для любых других практических расчетов, приходится обращаться к опыту. Методы анализа результатов опытов и определения по ним вероятностей событий и характеристик случайных величин дает математическая статистика — обширный раздел современной теории вероятностей.

Из нашего подхода к определению основных понятий теории вероятностей в предыдущих главах непосредственно вытекают и способы определения основных статистических характеристик из опыта. Так, мы знаем, что за вероятность события можно принять его частоту при большом числе опытов, за математическое ожидание и дисперсию случайной величины — выборочное среднее и выборочную дисперсию, за ковариацию и коэффициент корреляции двух случайных величин — выборочные ковариацию и коэффициент корреляции. Наконец, плотность случайной величины можно определить как относительную плотность экспериментальных точек в соответствующих интервалах. Однако при этом остается много неясных вопросов. При любом измерении или экспериментальном определении какой-либо величины всегда возникает вопрос о точности определения этой величины. Так и в нашем случае прежде всего возникает вопрос: с какой точностью мы получаем вероятность события, принимая за нее частоту этого события? Что надо сделать, чтобы повысить точность определения вероятности события? Сколько нужно произвести опытов, чтобы получить вероятность события с заданной точностью? Подобные вопросы возникают и при определении характеристик слу-

чайных величин. Все эти вопросы необходимо уметь решать, чтобы знать, насколько можно доверять полученным из опыта данным и основанным на них расчетам, полагаться на них в нашей практической деятельности.

Прежде всего, ясно, что не может быть речи о точном определении вероятности события по конечному числу опытов. Конечно, случайно частота события может оказаться равной его вероятности. Но это маловероятно и обнаружить это совпадение невозможно. Частота события, полученная в результате опытов, практически всегда отличается от его вероятности.

Однако и о приближенном в обычном смысле определении вероятности события по результатам опытов говорить не приходится. Говоря о приближенном определении какой-либо величины, мы обычно подразумеваем, что можно указать пределы погрешности, из которых ошибка не выйдет. Частота же события случайна при любом числе опытов из-за случайности результатов отдельных опытов. Так, например, если мы будем много раз определять частоту одного и того же события, скажем появления герба, по результатам 100 опытов, то будем получать различные значения частоты этого события. Из-за случайности результатов отдельных опытов частота может довольно значительно отклоняться от вероятности события. Поэтому, определяя неизвестную вероятность события как частоту этого события при большом числе опытов, мы не можем указать пределы погрешности и гарантировать, что ошибка не выйдет из этих пределов. Из-за случайности результатов опытов полной гарантии, что ошибка не выйдет из любых заданных пределов, быть не может. Поэтому в математической статистике обычно говорят не о приближенных значениях неизвестных величин, а об их подходящих значениях, оценках.

В дальнейшем мы будем называть все подлежащие определению величины — вероятности событий, числовые характеристики и распределения случайных величин — *статистическими характеристиками*, а найденные по результатам опытов их значения — *оценками*.

Оценив по результатам опытов ту или иную статистическую характеристику, естественно поставить вопрос: насколько согласуется с опытными данными предположение (гипотеза) о том, что неизвестная характеристика имеет именно то значение, которое получено в результате

ее оценивания? Так возникает второй важный класс задач математической статистики — задачи *проверки гипотез*.

Таким образом, основными задачами математической статистики являются разработка методов нахождения оценок и исследования точности их приближения к оцениваемым характеристикам и разработка методов проверки гипотез.

1.2. Виды вероятностной сходимости. В главах 1 и 3 на основе интуитивных представлений мы пришли к выводу, что при увеличении числа опытов n частоты событий и выборочные средние значений случайных величин должны приближаться к соответствующим вероятностям и математическим ожиданиям. Однако вследствие случайности результатов опытов о сходимости в обычном смысле здесь говорить не приходится. В связи с этим возникает необходимость введения новых понятий сходимости и предела, отличных от принятых в элементарном математическом анализе.

Последовательность случайных величин $\{S_n\}$ называется *сходящейся в среднем квадратическом (с. к. сходящейся)* к случайной величине S, если $M|S_n - S|^2 \to 0$ при $n \to \infty$.

Последовательность случайных величин $\{S_n\}$ называется *сходящейся по вероятности (п. в.)* к случайной величине S, если при любом $\varepsilon > 0$

$$P(|S_n - S| \geqslant \varepsilon) \to 0 \quad \text{при} \quad n \to \infty.$$

Последовательность случайных величин $\{S_n\}$ называется *сходящейся почти наверное (п. н.)* к случайной величине S, если $P(S_n \to S) = 1$.

Эти определения относятся как к скалярным, так и к конечномерным векторным случайным величинам. В последнем случае $|x|$ означает модуль (норму) вектора x.

Так как любую неслучайную величину можно рассматривать как случайную величину с единственным возможным значением, имеющим вероятность, равную единице, то приведенные определения относятся и к сходимости последовательности случайных величин к неслучайной величине, в частности к нулю.

1.3. Неравенство Чебышева. Связь между разными видами сходимости. Интуитивно ясно, что *всякая с. к. сходящаяся последовательность сходится к тому же пределу и п. в.* Для доказательства выведем сначала одно важное в теории вероятностей неравенство.

▶ При произвольных $\varepsilon > 0$ и $p > 0$ для любой случайной величины X (скалярной или векторной) имеем

$$P(|X| \geqslant \varepsilon) = \int\limits_{|x| \geqslant \varepsilon} f(x)\,dx \leqslant \frac{1}{\varepsilon^p} \int\limits_{|x| \geqslant \varepsilon} |x|^p f(x)\,dx \leqslant$$

$$\leqslant \frac{1}{\varepsilon^p} \int\limits_{-\infty}^{\infty} |x|^p f(x)\,dx = \frac{1}{\varepsilon^p} M|X|^p.$$

Здесь в первом и втором интегралах неравенство под знаком интеграла означает, что областью интегрирования служит множество всех x, удовлетворяющих этому неравенству.

Таким образом, при любых $\varepsilon > 0$, $p > 0$

$$P(|X| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^p} M|X|^p. \blacktriangleleft \tag{1}$$

Это неравенство обычно называется *неравенством Чебышева*. При практическом применении этого неравенства число $p > 0$ выбирается так, чтобы было $M|X|^p < \infty$.

Применив неравенство (1) к случайной величине $X = S_n - S$ при $p = 2$, получаем

$$P(|S_n - S| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} M|S_n - S|^2.$$

Отсюда непосредственно следует сходимость п. в. всякой с. к. сходящейся последовательности.

Обратное в общем случае неверно. Чтобы доказать это, достаточно привести соответствующий пример.

Пример 1. Пусть S_n — случайная величина с плотностью

$$f_n(s) = \frac{1}{\pi} \frac{n}{1 + n^2 s^2} \qquad (n = 1, 2, \ldots)$$

(распределение Коши). Так как

$$P(|S_n| < \varepsilon) = \int\limits_{-\varepsilon}^{\varepsilon} f_n(s)\,ds = \frac{1}{\pi} \int\limits_{-n\varepsilon}^{n\varepsilon} \frac{dt}{1 + t^2} = \frac{2}{\pi} \operatorname{arctg}(n\varepsilon) \longrightarrow 1$$

при $n \to \infty$, то $P(|S_n| \geqslant \varepsilon) \to 0$ при любом $\varepsilon > 0$. Поэтому последовательность $\{S_n\}$ сходится п. в. к нулю. Однако $M|S_n|^2 = \infty$ при любом n (пример 3.6), вследствие чего последовательность $\{S_n\}$ не сходится с. к. к нулю.

Легко сообразить, что и *всякая последовательность случайных величин $\{S_n\}$, сходящаяся п. н. к S, сходится к S и п. в.*

▶ Действительно, из сходимости п. н. $\{S_n\}$ к S следует, что для любого $\varepsilon > 0$

$$P(|S_m - S| \geqslant \varepsilon \text{ хотя бы при одном } m \geqslant n) \longrightarrow 0 \text{ при } n \longrightarrow \infty.$$

Но

$$P(|S_n - S| \geqslant \varepsilon) \leqslant \leqslant P(|S_m - S| \geqslant \varepsilon \text{ хотя бы при одном } m \geqslant n),$$

откуда и следует сходимость п. в. последовательности $\{S_n\}$ к S. ◀

Обратное в общем случае неверно.

Пример 2*). Пусть S_n — случайная величина с двумя возможными значениями 0 и 1, вероятности которых равны соответственно $1-(1/n)$ и $1/n$, причем при любых натуральных $n_1, \ldots, n_k$ случайные величины $S_{n_1}, \ldots, S_{n_k}$ независимы. Последовательность $\{S_n\}$ сходится п. в. к 0, так как при любом ε, $0 < \varepsilon < 1$,

$$P(|S_n| > \varepsilon) = P(S_n = 1) = \frac{1}{n} \longrightarrow 0 \quad \text{при} \quad n \longrightarrow \infty.$$

Однако эта последовательность не сходится п. н. к 0. Действительно, при любом ε, $0 < \varepsilon < 1$,

$$P(|S_m| < \varepsilon \text{ при всех } m \geqslant n) \leqslant \leqslant P(|S_m| < \varepsilon \text{ при всех } m,\ n \leqslant m \leqslant N) = = \prod_{m=n}^{N} P(S_m = 0) = \frac{n-1}{n}\,\frac{n}{n+1} \cdots \frac{N-2}{N-1}\,\frac{N-1}{N} = \frac{n-1}{N}.$$

Отсюда в силу произвольности N следует, что при любом n вероятность $P(|S_m| < \varepsilon$ при всех $m \geqslant n)$ меньше сколь угодно малого положительного числа. А это и значит, что вероятность сходимости последовательности $\{S_n\}$ равна 0, т. е. $\{S_n\}$ не сходится п. н. к нулю.

Заметим, что последовательность $\{S_n\}$ в данном случае с. к. сходится к нулю, так как $M|S_n|^2 = 1/n \longrightarrow 0$ при $n \longrightarrow \infty$. Таким образом, мы имеем пример последовательности случайных величин, сходящейся п. в. и с. к., но не сходящейся п. н.

Пример 3. Пусть S_n — случайная величина с двумя возможными значениями 0 и n, вероятности которых равны соответственно $1-(1/n^2)$ и $1/n^2$. Последовательность $\{S_n\}$ сходится п. в. к нулю, так как при любом $\varepsilon > 0$ для всех достаточно больших n

$$P(|S_n| \geqslant \varepsilon) = P(S_n = n) = \frac{1}{n^2} \longrightarrow 0 \quad \text{при} \quad n \longrightarrow \infty.$$

Последовательность $\{S_n\}$ сходится к 0 и п. н., так как

$$P(|S_m| \geqslant \varepsilon \text{ хотя бы при одном } m \geqslant n) \leqslant \leqslant \sum_{m=n}^{\infty} P(S_m = m) = \sum_{m=n}^{\infty} \frac{1}{m^2} \longrightarrow 0 \quad \text{при} \quad n \longrightarrow \infty$$

*) Примеры 2 и 3 заимствованы из книги Бартлетта [3].

и, следовательно, $P(|S_m| < \varepsilon$ при всех $m \geqslant n) \longrightarrow 1$ при $n \longrightarrow \infty$, что и доказывает сходимость последовательности $\{S_n\}$ к нулю с вероятностью 1, т. е. сходимость п. н.

Однако последовательность $\{S_n\}$ не сходится с. к. к нулю, так как $M|S_n|^2 = 1$ при любом n.

Пример 4. Пусть S_n — случайная величина с возможными значениями 0 и 1, вероятности которых равны соответственно $1-(1/n^2)$ и $1/n^2$. Показать, что последовательность $\{S_n\}$ сходится к 0 и п. в., и п. н., и с. к.

Приведенные примеры показывают, что последовательность случайных величин может быть сходящейся п. в., но не сходящейся с. к. или п. н., может быть сходящейся п. н., но не сходящейся с. к. и наоборот. Таким образом, как класс с. к. сходящихся последовательностей, так и класс п. н. сходящихся последовательностей являются подклассами последовательностей, сходящихся п. в., причем эти два подкласса имеют непустое пересечение.

Так как в силу (3.29)

$$M|S_n - S|^2 = D(S_n - S) + |MS_n - MS|^2, \tag{2}$$

то *для с. к. сходимости последовательности* $\{S_n\}$ *к* S *необходима и достаточна сходимость последовательности математических ожиданий* $\{MS_n\}$ *к* MS *и сходимость последовательности дисперсий разностей* $S_n - S$ *к нулю.*

Из (2) следует также, что для с. к. сходимости последовательности случайных величин $\{S_n\}$ к неслучайной величине a необходима и достаточна сходимость последовательности математических ожиданий $\{MS_n\}$ к a и сходимость последовательности дисперсий $\{DS_n\}$ к нулю.

Сходимость по вероятности, сходимость в среднем квадратическом и сходимость почти наверное представляют собой особые, вероятностные виды сходимости. В частном случае последовательности неслучайных величин все эти три вида сходимости совпадают с обычной сходимостью. Таким образом, каждый из трех видов вероятностной сходимости представляет собой естественное обобщение обычного понятия сходимости на случайные величины.

§ 2. Оценивание статистических характеристик

2.1. Оценки. В математической статистике обычно рассматривают только действительные случайные величины. Это не ограничивает общности, так как любую комплекс-

ную случайную величину можно рассматривать как двумерный случайный вектор с действительными координатами. Поэтому везде в этой главе мы будем считать все случайные величины, в том числе и функции случайных величин, действительными.

Результаты опытов мы будем везде рассматривать как случайные величины. Поэтому будем обозначать результаты опытов и все функции результатов опытов большими буквами. При практическом применении полученных формул следует заменять в них все случайные величины их реализациями, полученными в результате опытов.

Любая функция результатов опытов, которая не зависит от неизвестных статистических характеристик, называется *статистикой*.

Оценкой статистической характеристики θ называется статистика, реализация которой, полученная в результате опытов, принимается за неизвестное истинное значение параметра θ.

Ясно, что не всякая статистика может служить оценкой данной статистической характеристики. Поскольку результаты опытов случайны, любая статистика представляет собой случайную величину. Чтобы статистика могла служить оценкой данной статистической характеристики θ, необходимо, чтобы распределение этой статистики было сосредоточено в достаточной близости от неизвестного значения θ, так, чтобы вероятность больших отклонений этой статистики от θ была достаточно мала. Тогда при систематическом многократном применении этой статистики в качестве оценки данной характеристики в среднем будет получаться достаточная точность. Вероятность больших отклонений будет мала, и они будут встречаться редко. Желательно еще, чтобы при увеличении числа опытов точность результатов оценивания увеличивалась. В связи с этим вводят следующие определения, касающиеся оценок.

Оценка статистической характеристики θ называется *состоятельной*, если она сходится по вероятности к θ при неограниченном увеличении числа опытов n.

Чтобы оценка характеристики θ была состоятельной, достаточно, чтобы ее математическое ожидание стремилось к θ, а ее дисперсия стремилась к нулю при неограниченном увеличении числа опытов n. Это непосредственно следует из результатов п. 1.3.

Оценка $\hat{\Theta}$ статистической характеристики θ называется *несмещенной*, если ее математическое ожидание равно θ при любом числе опытов n. Оценка $\hat{\Theta}$ характеристики θ называется *смещенной*, если ее математическое ожидание не равно θ. Разность $M\hat{\Theta}-\theta$ называется смещением оценки $\hat{\Theta}$.

Качество оценки часто характеризуют математическим ожиданием квадрата модуля отклонения оценки от оцениваемой характеристики (средним квадратом ошибки):

$$\delta = M\,|\hat{\Theta}-\theta|^2.$$

В случае несмещенной оценки $\hat{\Theta}$ средний квадрат ошибки δ представляет собой дисперсию оценки $\hat{\Theta}$ в случае скалярной характеристики θ и сумму дисперсий координат оценки $\hat{\Theta}$ в случае векторной характеристики θ.

Множество чисел δ, соответствующих всем оценкам $\hat{\Theta}$ данной характеристики θ, при данном числе опытов n, как и всякое множество неотрицательных чисел, имеет точную нижнюю грань

$$\delta_0 = \inf_{\hat{\Theta}} M\,|\hat{\Theta}-\theta|^2.$$

Естественно стремление пользоваться такими оценками, для которых средний квадрат ошибки δ или равен δ_0, или близок к δ_0. Для нахождения таких оценок во многих случаях оказывается полезным понятие достаточной статистики.

2.2. Достаточные статистики. Статистика S (скалярная или векторная) называется *достаточной* для характеристики θ, если при любом выборе другой статистики S' (тоже скалярной или векторной), для которой не существует зависимостей вида $\psi(S, S')=0$, условное распределение статистики S' при данном значении s статистики S не зависит от θ. Иными словами, статистика S достаточна, если знание значений любых других статистик не дает никакой дополнительной информации о θ сверх той, которая содержится в S.

Если достаточная статистика существует, то она не может быть определена однозначно.

▶ Действительно, предположим, что статистика S достаточна. На основании (4.16) условная плотность любой статистики S' при данном значении s величины S

определяется формулой

$$f_2(s' \mid s) = \frac{f(s, s' \mid \theta)}{f_1(s \mid \theta)}, \tag{3}$$

где $f(s, s' \mid \theta)$ — совместная плотность величин S, S', зависящая от параметра θ, а $f_1(s \mid \theta)$ — плотность величины S. Левая часть формулы (3) не зависит от θ вследствие достаточности статистики S. Рассмотрим любую функцию $S_1 = \varphi(S)$ статистики S, такую, что уравнение $s_1 = \varphi(s)$ имеет единственное решение $s = \varphi^{-1}(s_1)$ в области возможных значений величины S. Пользуясь формулой (5.31), находим плотность величины S_1 и совместную плотность величин S_1, S':

$$g_1(s_1 \mid \theta) = f_1(\varphi^{-1}(s_1) \mid \theta) \mid J(s_1) \mid,$$
$$g(s_1, s' \mid \theta) = f(\varphi^{-1}(s_1), s' \mid \theta) \mid J(s_1) \mid,$$

где $J(s_1)$ — якобиан координат функции $\varphi^{-1}(s_1)$ по координатам вектора s_1. Условная плотность статистики S' при данном значении s_1 статистики S_1 согласно (4.16) определяется формулой

$$g_2(s' \mid s_1) = \frac{g(s_1, s' \mid \theta)}{g_1(s_1 \mid \theta)} = \frac{f(\varphi^{-1}(s_1), s' \mid \theta)}{f_1(\varphi^{-1}(s_1) \mid \theta)}.$$

Сравнив эту формулу с (3), получаем

$$g_2(s' \mid s_1) = f_2(s' \mid \varphi^{-1}(s_1)).$$

Так как правая часть этой формулы не зависит от θ, то и левая часть не зависит от θ, что и доказывает достаточность статистики $S_1 = \varphi(S)$. ◀

Таким образом, *любая статистика, связанная с достаточной статистикой взаимно однозначной зависимостью, тоже достаточна.*

Докажем, что *если совместная плотность результатов опытов $X_1, \ldots, X_n$ может быть представлена в виде*

$$g(x_1, \ldots, x_n \mid \theta) = h(\varphi(x_1, \ldots, x_n) \mid \theta),$$

то статистика $S = \varphi(X_1, \ldots, X_n)$ достаточна.

▶ Пусть $S' = \psi(X_1, \ldots, X_n)$ — любая статистика, для которой не существует зависимостей вида $\chi(S, S') = 0$. Предположим сначала, что сумма размерностей векторов S и S' равна сумме размерностей векторов $X_1, \ldots, X_n$ и что при любых значениях s и s' из областей значений функций φ и ψ уравнения

$$\varphi(x_1, \ldots, x_n) = s, \quad \psi(x_1, \ldots, x_n) = s'$$

имеют множество решений $\{x_k^{(l)}(s, s'),\ k=1, \ldots, n\}$, $l \in I(s, s')$. Тогда совместная плотность величин S и S' определится по формуле (5.34):

$$f(s, s' \mid \theta) = \sum_{l \in I(s, s')} g(x_1^{(l)}(s, s'), \ldots, x_n^{(l)}(s, s') \mid \theta) \mid J_l(s, s') \mid,$$

где $J_l(s, s')$ — якобиан координат векторных функций $x_1^{(l)}(s, s'), \ldots, x_n^{(l)}(s, s')$ по координатам векторов s и s', $l \in I(s, s')$. Но

$$g(x_1^{(l)}(s, s'), \ldots, x_n^{(l)}(s, s') \mid \theta) = \\ = h(\varphi(x_1^{(l)}(s, s'), \ldots, x_n^{(l)}(s, s')) \mid \theta) = h(s \mid \theta),$$

так как функции $x_1 = x_1^{(l)}(s, s'), \ldots, x_n = x_n^{(l)}(s, s')$, $l \in I(s, s')$, удовлетворяют уравнениям

$$\varphi(x_1, \ldots, x_n) = s, \quad \psi(x_1, \ldots, x_n) = s'.$$

Следовательно,

$$f(s, s' \mid \theta) = h(s \mid \theta) \sum_{l \in I(s, s')} \mid J_l(s, s') \mid.$$

Отсюда, интегрируя по s', находим плотность величины S:

$$f_1(s \mid \theta) = h(s \mid \theta) \int \sum_{l \in I(s, s')} \mid J_l(s, s') \mid ds'.$$

Из полученных формул следует, что условная плотность величины S' при данном s

$$f_2(s' \mid s) = f(s, s' \mid \theta)/f_1(s \mid \theta)$$

не зависит от θ, что и доказывает достаточность статистики S.

Точно так же доказывается достаточность статистики $S = \varphi(X_1, \ldots, X_n)$ в случае, когда сумма размерностей векторов S и S' не совпадает с суммой размерностей векторов $X_1, \ldots, X_n$. ◀

2.3. Доверительные интервалы и области. Как уже было сказано в п. 1.1, вследствие случайности результатов опытов невозможно установить достаточно узкие пределы, из которых ошибка оценки (т. е. отклонение оценки от оцениваемой характеристики) не выходила бы с полной гарантией. Поэтому возникает задача определения по результатам опытов таких пределов, из которых ошибка оценки не выходила бы с заданной вероятностью. Например, приняв частоту события $\hat{P}$ за оценку его вероятности p, установить по результатам тех же опытов

такую границу возможных отклонений $\hat{P}$ от p, которую модуль ошибки $|\hat{P}-p|$ не превосходил бы с заданной вероятностью α. Эта граница, естественно, будет тоже случайной в силу случайности результатов опытов. Таким образом, речь идет о нахождении по результатам опытов такого случайного интервала (т. е. интервала со случайными концами), который с заданной вероятностью α содержал бы неизвестное значение вероятности p.

Случайный интервал, полностью определяемый результатами опытов и не зависящий от неизвестных характеристик, который с заданной вероятностью α накрывает неизвестную скалярную статистическую характеристику θ, называется *доверительным интервалом* для этой характеристики, соответствующим *коэффициенту доверия* α. Величина $1-\alpha$ называется *уровнем значимости* отклонения оценки. Концы доверительного интервала называются *доверительными границами*.

Обобщение понятия доверительного интервала на случай векторной характеристики θ приводит к понятию доверительной области.

Доверительной областью для векторной характеристики θ, соответствующей *коэффициенту доверия* α, называется такая случайная область, полностью определяемая результатами опытов и не зависящая от неизвестных характеристик, которая с вероятностью α содержит неизвестное значение θ.

Очевидно, что доверительная область определяется заданным коэффициентом доверия α не однозначно, существует бесконечное множество доверительных областей, соответствующих одному и тому же значению α. Обычно стараются определить доверительные области, имеющие минимальные размеры при данной вероятности α. Этому условию часто удовлетворяют хотя бы приблизительно области, симметричные относительно оценки $\hat{\Theta}$ характеристики θ.

2.4. Методы определения доверительных областей. Для определения доверительных интервалов и областей обычно применяются три основных метода.

Первый метод, удобный в случае положительного скалярного параметра θ, основан на нахождении распределения отношения оценки $\hat{\Theta}$ параметра θ к самому параметру θ. Если оценка $\hat{\Theta}$ такова, что это распределение не зависит от неизвестных характеристик, то, зная это

распределение, можно найти вероятность попадания отношения $\hat{\Theta}/\theta$ в любой интервал и, наоборот, по заданной вероятности α найти интервал, вероятность попадания $\hat{\Theta}/\theta$ в который равна α. Любой такой интервал будет доверительным интервалом для θ. Обычно стараются получить доверительный интервал, симметричный относительно оценки. Однако это не всегда возможно. В некоторых случаях нижняя доверительная граница получается отрицательной, что не имеет смысла. Поэтому доверительный интервал для положительного параметра θ определяют формулой

$$\max\{0, (1-\varepsilon_\alpha)\hat{\Theta}\} < \theta < (1+\varepsilon_\alpha)\hat{\Theta}. \tag{4}$$

При $\varepsilon_\alpha \in (0, 1]$ этот интервал симметричен относительно $\hat{\Theta}$. При $\varepsilon_\alpha > 1$ симметрия не достигается. Неравенства (4) выполняются тогда и только тогда, когда

$$\frac{1}{1+\varepsilon_\alpha} < \frac{\hat{\Theta}}{\theta} < \frac{1}{\max\{0, 1-\varepsilon_\alpha\}}.$$

Поэтому формула (4) определяет доверительный интервал для θ, соответствующий коэффициенту доверия α, если ε_α удовлетворяет уравнению

$$P\left(\frac{1}{1+\varepsilon_\alpha} < \frac{\hat{\Theta}}{\theta} < \frac{1}{\max\{0, 1-\varepsilon_\alpha\}}\right) = \alpha.$$

Второй метод состоит в том, что для каждого возможного значения неизвестного параметра θ выбирают такую область, содержащую θ, в которую оценка $\hat{\Theta}$ попадает с заданной вероятностью α. Эта область в общем случае зависит от θ и, конечно, от α. Поэтому мы обозначим ее $D_\alpha(\theta)$. Тогда можем написать

$$P(\hat{\Theta} \in D_\alpha(\theta)) = \alpha.$$

После этого для каждого фиксированного значения оценки $\hat{\Theta}$ определяют множество значений параметра θ, при которых $\hat{\Theta} \in D_\alpha(\theta)$. Это множество $\{\theta: \hat{\Theta} \in D_\alpha(\theta)\}$ в общем случае зависит от $\hat{\Theta}$ и α. Поэтому мы обозначим его $\Delta_\alpha(\hat{\Theta})$. Очевидно, что $\theta \in \Delta_\alpha(\hat{\Theta})$ тогда и только тогда, когда $\hat{\Theta} \in D_\alpha(\theta)$ при этом θ. Действительно, если при данном значении θ оценка $\hat{\Theta}$ приняла значение $\hat{\theta} \in D_\alpha(\theta)$, то по определению области $\Delta_\alpha(\hat{\theta})$ данное значение θ при-

надлежит $\Delta_\alpha(\hat{\theta})$, $\theta \in \{\theta: \hat{\theta} \in D_\alpha(\theta)\} = \Delta_\alpha(\hat{\theta})$. Наоборот, если оценка $\hat{\Theta}$ приняла такое значение $\hat{\theta}$, что $\theta \in \Delta_\alpha(\hat{\theta}) = \{\theta: \hat{\theta} \in D_\alpha(\theta)\}$, то $\hat{\theta} \in D_\alpha(\theta)$. Следовательно, при любом θ

$$P(\theta \in \Delta_\alpha(\hat{\Theta})) = P(\hat{\Theta} \in D_\alpha(\theta)) = \alpha. \tag{5}$$

Это равенство показывает, что область $\Delta_\alpha(\hat{\Theta})$ представляет собой доверительную область для параметра θ, соответствующую доверительному уровню α.

Третий метод основан на нахождении скалярной функции $\varphi(\hat{\Theta}, S, \theta)$ оценки $\hat{\Theta}$, некоторой другой статистики S и неизвестного параметра θ, обладающей следующими свойствами:

1) при любом значении s статистики S и любом значении θ неравенство $\varphi(\hat{\theta}, s, \theta) < c$, $c > 0$, при возрастании c определяет монотонно возрастающее семейство вложенных одна в другую областей $D(s, \theta, c) = \{\hat{\theta}: \varphi(\hat{\theta}, s, \theta) < c\}$;

2) $\varphi(\theta, s, \theta) = 0$ при любых s, θ и $\varphi(\hat{\theta}, s, \theta) > 0$ при любых $\hat{\theta}$, s, θ, $\hat{\theta} \neq \theta$, и, следовательно, точка $\hat{\theta} = \theta$ принадлежит области $D(s, \theta, c)$ при любых $c > 0$, s, θ;

3) распределение случайной величины $T = \varphi(\hat{\Theta}, S, \theta)$ не зависит от θ. Зная распределение величины $T = \varphi(\hat{\Theta}, S, \theta)$, можно определить такое $\varepsilon_\alpha > 0$, чтобы с вероятностью α выполнялось неравенство $\varphi(\hat{\Theta}, S, \theta) < \varepsilon_\alpha$:

$$P(\varphi(\hat{\Theta}, S, \theta) < \varepsilon_\alpha) = \alpha. \tag{6}$$

Эта формула определяет случайную область $\varphi(\hat{\Theta}, S, \theta) < \varepsilon_\alpha$, накрывающую неизвестное значение параметра θ с вероятностью α, т. е. доверительную область для θ, соответствующую коэффициенту доверия α.

Многомерные доверительные области трудно представить наглядно. Поэтому возникает вопрос о нахождении таких доверительных интервалов $I_1(\hat{\Theta}), \ldots, I_r(\hat{\Theta})$ для координат $\theta_1, \ldots, \theta_r$ векторного параметра θ, чтобы вероятность одновременного накрытия всех $\theta_1, \ldots, \theta_r$ соответствующими интервалами была не меньше заданного α (т. е. о нахождении прямоугольной доверительной области для θ, соответствующей коэффициенту доверия, не меньшему чем α).

Введем события $A_k = \{\theta_k \in I_k(\hat{\Theta})\}$ $(k = 1, \ldots, r)$ и обозначим прямоугольную область в r-мерном пространстве,

образованную интервалами $I_1(\hat{\Theta}), \ldots, I_r(\hat{\Theta})$, через $I(\hat{\Theta})$. На основании принципа двойственности и свойства полуаддитивности вероятности (1.13)

$$P(\theta \in I(\hat{\Theta})) = P\left(\bigcap_{k=1}^{r} A_k\right) = 1 - P\left(\bigcup_{k=1}^{r} \bar{A}_k\right) \geqslant 1 - \sum_{k=1}^{r} P(\bar{A}_k).$$

Отсюда видно, что для решения поставленной задачи достаточно найти доверительные интервалы $I_1(\hat{\Theta}), \ldots$ $\ldots, I_r(\hat{\Theta})$ для координат вектора θ, соответствующие коэффициенту доверия $1-(1-\alpha)/r$. Тогда вероятность одновременного накрытия этими интервалами всех $\theta_1, \ldots, \theta_r$ будет не меньше α. Фактически она во многих случаях может быть значительно больше α*).

Основную трудность в практическом применении изложенных методов представляет нахождение распределений различных статистик, в частности оценок. В настоящее время эти вопросы достаточно хорошо разработаны только для случая нормального распределения наблюдаемых случайных величин.

При определении оценок статистических характеристик и доверительных областей для них по результатам опытов обычно считают опыты независимыми в смысле определения, данного в § 1.8. В соответствии с этим определением будем считать опыты независимыми, если вероятность любого события и распределение любой случайной величины в каждом опыте не зависят от результатов предшествующих опытов. Кроме того, предположим, что вероятности наблюдаемых событий и распределения наблюдаемых случайных величин не изменяются от опыта к опыту. Таким образом, везде в дальнейшем будем считать случайные результаты различных опытов независимыми событиями и случайными величинами, имеющими одни и те же вероятности и распределения во всех опытах.

Нахождение оценок статистических характеристик обычно называют *точечным оцениванием*, а нахождение доверительных интервалов — *интервальным оцениванием*.

---

*) Другие методы нахождения совместных доверительных интервалов для координат векторного параметра читатель найдет в [103].

§ 3. Частота как оценка вероятности

3.1. Состоятельность. Пусть p — неизвестная вероятность интересующего нас события A, $\hat{P}$ — частота этого события при n опытах, рассматриваемая как случайная величина — функция случайных результатов опытов.

▶ Чтобы исследовать частоту $\hat{P}$ как оценку вероятности события, воспользуемся формулами для математического ожидания и дисперсии частоты, полученными в примере 3.11:

$$M\hat{P}=p, \quad D\hat{P}=\frac{pq}{n}, \quad q=1-p. \tag{7}$$

Отсюда видно, что математическое ожидание частоты события равно его вероятности, а дисперсия частоты стремится к нулю при неограниченном увеличении числа опытов n. Следовательно, частота события с.к. сходится к его вероятности при $n \to \infty$. Но из с.к. сходимости вытекает и сходимость по вероятности (п.в.).

Таким образом, *частота события сходится п.в. к вероятности этого события при неограниченном увеличении числа опытов n.* В этом состоит теорема Якова Бернулли — первая и простейшая теорема из серии теорем, объединяемых под названием *закона больших чисел*, который устанавливает вероятностную сходимость экспериментальных статистических характеристик к соответствующим теоретическим характеристикам. Таким образом, закон больших чисел подтверждает наши интуитивные представления об устойчивости частот событий и характеристик случайных величин и тем самым обосновывает все практические применения теории вероятностей.

На основании доказанной теоремы Якова Бернулли и первой формулы (7) *частота события является состоятельной несмещенной оценкой его вероятности.* Более того, как мы увидим в § 7.1, частота представляет собой несмещенную оценку вероятности с минимальной дисперсией. Поэтому никакие другие оценки вероятности обычно не применяются. ◀

Остановимся несколько подробнее на смысле выводов, вытекающих из проведенного исследования. Из теоремы Якова Бернулли следует, что при любых $\varepsilon, \delta > 0$ при всех достаточно больших числах опытов n имеет место неравенство

$$P(|\hat{P}-p| \geqslant \varepsilon) < \delta. \tag{8}$$

В силу произвольности ε и δ и принципа практической уверенности отсюда следует, что при достаточно большом числе опытов n получение отклонения частоты от вероятности, превосходящего по абсолютной величине произвольно малое наперед заданное число ε, можно считать невозможным событием. На этом и основано использование частоты в качестве оценки вероятности события.

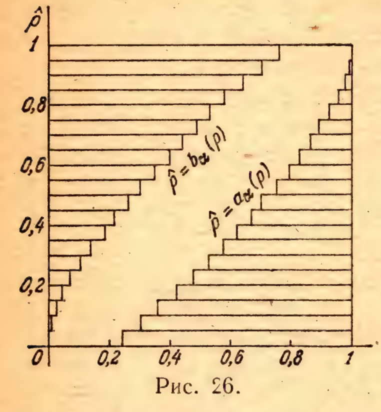


Рис. 26.

3.2. Доверительные интервалы. Для нахождения доверительных интервалов для вероятности обычно пользуются вторым методом п. 2.4.

▶ Для каждого значения вероятности $p \in (0, 1)$ задают интервал $D_\alpha(p) = [a_\alpha(p), b_\alpha(p))$, концы которого определяются из условий

$$P(\hat{P} < a_\alpha(p)) = F(a_\alpha(p)) \leqslant 1 - \frac{\alpha}{2},$$
$$P(\hat{P} \geqslant b_\alpha(p)) = 1 - F(b_\alpha(p)) \leqslant 1 - \frac{\alpha}{2}, \qquad (9)$$

где $F(x)$ — функция распределения частоты, т. е. функция биномиального распределения

$$F(x) = \sum_{m=0}^{n} C_n^m p^m q^{n-m} 1\left(x - \frac{m}{n}\right).$$

Тогда будем иметь

$$P(\hat{P} \in D_\alpha(p)) = P(a_\alpha(p) \leqslant \hat{P} < b_\alpha(p)) \geqslant \alpha.$$

Знак $\geqslant$ вместо $=$ здесь и соответственно $\leqslant$ в (9) берется потому, что вследствие дискретности частоты точные равенства, как правило, недостижимы. Для определения соответствующего доверительного интервала $\Delta_\alpha(\hat{P}) = \{p: a_\alpha(p) \leqslant \hat{P} < b_\alpha(p)\}$ достаточно найти его границы путем решения неравенств $a_\alpha(p) \leqslant \hat{P}$ и $\hat{P} < b_\alpha(p)$ относительно p. Вместо этого обычно пользуются следующим графическим построением. Определив $a_\alpha(p)$ и $b_\alpha(p)$ для каждого $p \in (0, 1)$, строят в координатах $(p, \hat{p})$ кривые $\hat{p} = a_\alpha(p)$ и $\hat{p} = b_\alpha(p)$ (рис. 26). Эти кривые при

каждом значении p определяют на соответствующей вертикальной прямой интервал $D_\alpha(p) = [a_\alpha(p),\ b_\alpha(p))$. Очевидно, что, и наоборот, при каждом значении $\hat{p}$ частоты $\hat{P}$ эти кривые определяют на соответствующей горизонтальной прямой интервал $\Delta_\alpha(\hat{p}) = \{p\colon \hat{p} \in [a_\alpha(p),\ b_\alpha(p))\}$, представляющий собой реализацию доверительного интервала $\Delta_\alpha(\hat{P})$, соответствующую данной реализации $\hat{p}$ частоты $\hat{P}$ (рис. 26). ◀

В силу дискретности частоты кривые $\hat{p} = a_\alpha(p)$, $\hat{p} = b_\alpha(p)$ являются ступенчатыми, причем высоты всех ступенек равны $1/n$, а ширина каждой ступеньки равна длине соответствующего интервала значений p, при которых $a_\alpha(p)$ сохраняет постоянное значение до очередного изменения скачком на $1/n$. Однако для простоты их обычно заменяют непрерывными кривыми. Графики для определения доверительных интервалов при оценивании вероятностей для $\alpha = 0{,}95$ и $\alpha = 0{,}99$ были построены Клоппером и Е. Пирсоном [37].

3.3. Приближенное определение доверительных интервалов. При большом числе опытов n (практически при $n \geqslant 100$) определение доверительных интервалов для p существенно упрощается. Как показано в примере 5.46, распределение случайной величины $Y = \sqrt{n}\,(\hat{P} - p)/\sqrt{pq} = (X - np)/\sqrt{npq}$ стремится к нормальному распределению $N(0, 1)$ при $n \to \infty$. Вследствие этого биномиальное распределение при достаточно большом числе опытов n как угодно мало отличается от нормального. Это дает возможность при больших n пользоваться нормальной функцией распределения вместо точной биномиальной для приближенного определения доверительных интервалов для вероятности.

▶ Предположив, что частота $\hat{P}$ распределена нормально, в силу симметрии нормального распределения и формул (7) получим

$$a_\alpha(p) = p - \varepsilon_\alpha \sqrt{pq/n}, \quad b_\alpha(p) = p + \varepsilon_\alpha \sqrt{pq/n},$$

где ε_α определяется в силу (3.93) уравнением

$$P\left(|\hat{P} - p| < \varepsilon_\alpha \sqrt{\frac{pq}{n}}\right) = 2\Phi(\varepsilon_\alpha) = \alpha. \qquad (10)$$

Практически корень ε_α уравнения (10) определяется по

таблице 3*). Поскольку ε_α не зависит от p, кривые $\hat{p}=a_\alpha(p)$ и $\hat{p}=b_\alpha(p)$ в этом случае представляют собой части эллипса $n(\hat{p}-p)^2=\varepsilon_\alpha^2 p(1-p)$ с центром в точке $p=\hat{p}=1/2$, касающегося вертикальных прямых $p=0$ и $p=1$ (рис. 27). Доверительные границы P_1, P_2 в этом случае легко находятся из уравнения эллипса путем решения его относительно p:

$$P_1,\ P_2=\frac{\hat{P}+\varepsilon_\alpha^2/2n}{1+\varepsilon_\alpha^2/n}\pm\frac{\varepsilon_\alpha}{1+\varepsilon_\alpha^2/n}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}+\frac{\varepsilon_\alpha^2}{4n^2}}. \blacktriangleleft \quad (11)$$

Пример 5. При 100 опытах событие A появилось 68 раз. Таким образом, частота $\hat{P}$ события A приняла значение 0,68. Найдем доверительный интервал для вероятности $p=P(A)$, соответствующий коэффициенту доверия $\alpha=0{,}95$. Так как число опытов n в данном случае велико, то можно считать, что частота $\hat{P}$ распределена приблизительно нормально, и применить упрощенный метод. В соответствии с (10) находим по таблице 1 функции Лапласа $\Phi(u)$ значение ε_α аргумента u, при котором она равна $\alpha/2=0{,}475$; получаем $\varepsilon_\alpha=1{,}96$. После этого по формулам (11) находим $P_1\approx 0{,}583$, $P_2\approx 0{,}763$. Таким образом, доверительным интервалом для вероятности $p=P(A)$, соответствующим коэффициенту доверия $\alpha=0{,}95$, в данном случае служит интервал (0,583; 0,763). Этот интервал представляет собой реализацию случайного интервала (P_1, P_2), с вероятностью $\alpha=0{,}95$ содержащего неизвестную вероятность $p=P(A)$.

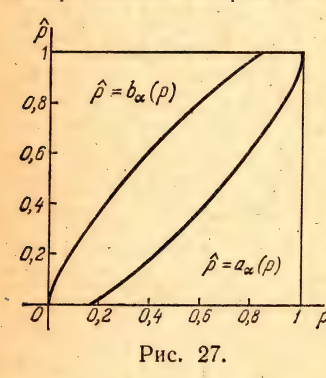


Рис. 27.

Вторая формула (7) показывает, что среднее квадратическое отклонение частоты обратно пропорционально $\sqrt{n}$. Следовательно, точность оценки вероятности повышается с увеличением числа опытов пропорционально $\sqrt{n}$. Как мы увидим дальше, эта закономерность распространяется и на моменты первого и второго порядков случайных величин.

*) При вычислениях на ЭВМ для определения ε_α можно пользоваться программой обращения нормальной функции распределения NDTRI [94].

§ 4. Оценки математического ожидания и дисперсии случайной величины

4.1. Оценка математического ожидания. Естественной оценкой математического ожидания случайной величины X служит среднее арифметическое ее экспериментальных значений $X_1, \ldots, X_n$ (выборочное среднее)

$$\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k. \qquad (12)$$

► Чтобы исследовать эту оценку, воспользуемся формулами для математического ожидания и дисперсии среднего арифметического, полученными в примере 3.12:

$$M\overline{X} = m_x, \qquad D\overline{X} = \frac{D_x}{n}. \qquad (13)$$

Отсюда следует, что выборочное среднее случайной величины с.к. сходится к ее математическому ожиданию при $n \to \infty$. Но из с.к. сходимости вытекает сходимость п.в.

Таким образом, мы доказали теорему Чебышева: *выборочное среднее случайной величины с конечными моментами первого и второго порядков сходится п.в. к ее математическому ожиданию при неограниченном увеличении числа опытов n.*

На основании теоремы Чебышева и первой формулы (13) *выборочное среднее представляет собой состоятельную несмещенную оценку математического ожидания случайной величины.* ◄

4.2. Оценки дисперсии. Естественной оценкой дисперсии случайной величины служит ее выборочная дисперсия

$$D_x^* = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})^2. \qquad (14)$$

► Для вычисления математического ожидания и дисперсии этой оценки, учитывая, что $MX_k = M\overline{X} = m_x$, представим (14) в виде

$$D_x^* = \frac{1}{n} \sum_{k=1}^{n} (X_k^0 - \overline{X}^0)^2 = \frac{1}{n} \left[\sum_{k=1}^{n} (X_k^0)^2 - 2\overline{X}^0 \sum_{k=1}^{n} X_k^0 + n(\overline{X}^0)^2 \right] = \frac{1}{n} \sum_{k=1}^{n} (X_k^0)^2 - (\overline{X}^0)^2.$$

Тогда будем иметь

$$MD_x^* = \frac{1}{n}\sum_{k=1}^{n} M(X_k^0)^2 - M(\overline{X}^0)^2 =$$

$$= \frac{1}{n}\sum_{k=1}^{n} DX_k - D\overline{X} = \frac{1}{n} nD_x - \frac{D_x}{n} = \frac{n-1}{n} D_x.$$

Таким образом,

$$MD_x^* = \frac{n-1}{n} D_x. \blacktriangleleft \tag{15}$$

Эта формула показывает, что D_x^* *является смещенной оценкой дисперсии* D_x *со смещением, равным* $-D_x/n$.

▶ Для вычисления дисперсии оценки D_x^* найдем сначала ее момент второго порядка MD_x^{*2}. Имея в виду, что

$$D_x^{*2} = \frac{1}{n^2}\left[\sum_{k=1}^{n}(X_k^0)^2\right]^2 - \frac{2}{n}(\overline{X}^0)^2 \sum_{k=1}^{n}(X_k^0)^2 + (\overline{X}^0)^4,$$

$$M\left[\sum_{k=1}^{n}(X_k^0)^2\right]^2 = M\sum_{k,\, l=1}^{n}(X_k^0)^2 (X_l^0)^2 =$$

$$= \sum_{k=1}^{n} M(X_k^0)^4 + \sum_{k\neq l} M(X_k^0)^2 M(X_l^0)^2 = n\mu_4 + n(n-1)\mu_2^2,$$

$$M\left[(\overline{X}^0)^2 \sum_{k=1}^{n}(X_k^0)^2\right] = \frac{1}{n^2} M \sum_{k,\, l,\, h=1}^{n} X_l^0 X_h^0 (X_k^0)^2 =$$

$$= \frac{1}{n^2}\left[\sum_{k=1}^{n} M(X_k^0)^4 + \sum_{k\neq l} M(X_k^0)^2 M(X_l^0)^2\right] = \frac{\mu_4 + (n-1)\mu_2^2}{n},$$

$$M(\overline{X}^0)^4 = \frac{1}{n^4} M\left[\sum_{k=1}^{n} X_k^0\right]^4 =$$

$$= \frac{1}{n^4}\left[\sum_{k=1}^{n} M(X_k^0)^4 + 3\sum_{k\neq l} M(X_k^0)^2 M(X_l^0)^2\right] = \frac{\mu_4 + 3(n-1)\mu_2^2}{n^3},$$

где $\mu_2 = D_x$, а μ_4 — четвертый центральный момент величины X, получаем после элементарных преобразований

$$MD_x^{*2} = \frac{n-2}{n^2}[\mu_4 + (n-1)\mu_2^2] + \frac{1}{n^3}[\mu_4 + 3(n-1)\mu_2^2].$$

После этого по формуле (3.29) находим

$$DD_x^* = MD_x^{*2} - (MD_x^*)^2 =$$

$$= \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}. \blacktriangleleft \tag{16}$$

При выводе этих формул мы воспользовались независимостью результатов опытов $X_1, \ldots, X_n$ и применили теорему умножения математических ожиданий (п. 4.2.6).

Формулы (15) и (16) показывают, что $MD_x^* \to D_x$, $DD_x^* \to 0$ при $n \to \infty$. Таким образом, *оценка* D_x^* *с.к. сходится к* D_x. Отсюда следует сходимость п.в. D_x^* к D_x, т. е. *состоятельность оценки* D_x^*.

▶ Чтобы получить несмещенную состоятельную оценку дисперсии D_x, достаточно, как это видно из (15), умножить оценку D_x^* на $n/(n-1)$. В результате получим оценку

$$\hat{D}_x = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2. \tag{17}$$

Дисперсия этой оценки в силу (16) определяется формулой

$$D\hat{D}_x = n\frac{\mu_4 - \mu_2^2}{(n-1)^2} - \frac{2(\mu_4 - 2\mu_2^2)}{(n-1)^2} + \frac{\mu_4 - 3\mu_2^2}{n(n-1)^2}. \tag{18}$$

В частном случае нормального распределения случайной величины X, как было показано в § 3.6, $\mu_4 = 3\mu_2^2$ и формула (18) принимает вид

$$D\hat{D}_x = \frac{2\mu_2^2}{n-1} = \frac{2D_x^2}{n-1}. \text{ ◀} \tag{19}$$

Для сравнения вычислим средний квадрат ошибки смещенной оценки D_x^* дисперсии D_x для случая нормального распределения X. В этом случае, учитывая, что $\mu_4 = 3\mu_2^2 = 3D_x^2$, из (15) и (16) получаем

$$M(D_x^* - D_x)^2 = [M(D_x^* - D_x)]^2 + DD_x^* = \frac{2n-1}{n^2}D_x^2.$$

Отсюда видно, что смещенная оценка D_x^* точнее несмещенной оценки $\hat{D}_x$, так как $(2n-1)/n^2 < 2/n < 2/(n-1)$.

4.3. Доверительные интервалы для математического ожидания. Задача определения доверительных интервалов для математического ожидания и дисперсии для произвольного числа опытов n решена только для случая нормально распределенной случайной величины X.

▶ Если дисперсия D_x известна, то доверительный интервал для математического ожидания находится очень просто. Для этого достаточно заметить, что среднее арифметическое $\overline{X}$ как линейная функция нормально распределенных случайных величин распределено нормально.

Поэтому на основании (13) и (3.93)

$$P\left(|\bar{X}-m_x|<\varepsilon_\alpha\sqrt{\frac{D_x}{n}}\right)=2\Phi(\varepsilon_\alpha).$$

Определив ε_α из условия $\Phi(\varepsilon_\alpha)=\alpha/2$, получим доверительный интервал для m_x, соответствующий коэффициенту доверия α, в виде

$$\bar{X}-\varepsilon_\alpha\sqrt{\frac{D_x}{n}}<m_x<\bar{X}+\varepsilon_\alpha\sqrt{\frac{D_x}{n}}. \blacktriangleleft \qquad (20)$$

При большом числе опытов n этот метод можно применить для приближенного определения доверительных интервалов для m_x при любом распределении величины X, так как при большом n распределение величины $\bar{X}$ стремится к нормальному (§ 5.4).

Оценивать математическое ожидание при известной дисперсии приходится при измерении какой-либо величины прибором, точность которого известна. В этом случае математическое ожидание результатов измерений равно сумме измеряемой величины и систематической ошибки прибора, а дисперсия равна известной дисперсии ошибки прибора.

► Для нахождения доверительного интервала для математического ожидания m_x в случае неизвестной дисперсии D_x обычно пользуются третьим методом п. 2.4. Введем n-мерный вектор результатов опытов $U=[X_1\dots X_n]^T$ и величину $S=U^TU-n\bar{X}^2$. Покажем, что случайная величина

$$T=\sqrt{n(n-1)}\,\frac{\bar{X}^0}{\sqrt{S}}$$

представляет собой стьюдентово отношение (пример 5.12). Для этого заметим, что при любом a

$$\sum_{k=1}^{n}(X_k-\bar{X})^2=\sum_{k=1}^{n}[X_k-a-(\bar{X}-a)]^2=$$

$$=\sum_{k=1}^{n}(X_k-a)^2-2(\bar{X}-a)\sum_{k=1}^{n}(X_k-a)+n(\bar{X}-a)^2=$$

$$=\sum_{k=1}^{n}(X_k-a)^2-n(\bar{X}-a)^2,$$

так как

$$\sum_{k=1}^{n}(X_k-a)=n(\bar{X}-a).$$

Положив в полученном равенстве один раз $a = m_x$, а другой раз $a = 0$, получим

$$\sum_{k=1}^{n} (X_k - \overline{X})^2 = \sum_{k=1}^{n} (X_k^0)^2 - n(\overline{X}^0)^2 = \sum_{k=1}^{n} X_k^2 - n\overline{X}^2,$$

откуда следует, что

$$(n-1)\hat{D}_x = \sum_{k=1}^{n} (X_k - \overline{X})^2 = U^{0T} U^0 - n(\overline{X}^0)^2 =$$
$$= U^T U - n\overline{X}^2 = S.$$

Это равенство показывает, что величина S выражается через центрированный вектор U^0 так же, как и в примере 5.12, что и доказывает наше утверждение. Следовательно, величина T имеет T-распределение с $k = n-1$ степенями свободы, определяемое плотностью

$$s_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{t^2}{2}\right)^{-\frac{k+1}{2}}.$$

Так как это распределение не зависит от неизвестных параметров m_x, D_x, то величина T может служить в качестве функции $\varphi(\hat{\Theta}, S, \theta)$ п. 2.4:

$$T = \varphi(\overline{X}, S, m_x) = \sqrt{n(n-1)}\,\frac{\overline{X} - m_x}{\sqrt{S}}.$$

Согласно п. 2.4 доверительный интервал для m_x определяется теперь неравенством

$$\left|\sqrt{n(n-1)}\,\frac{\overline{X} - m_x}{\sqrt{S}}\right| < \varepsilon_\alpha, \qquad (21)$$

где ε_α определяется уравнением

$$P(|T| < \varepsilon_\alpha) = S_k(\varepsilon_\alpha) = 2\int_0^{\varepsilon_\alpha} s_k(t)\,dt = \alpha. \qquad (22)$$

Практически ε_α находится непосредственно по таблице по данным k и α (табл. 3).

Чтобы преобразовать полученный доверительный интервал для m_x к более удобному виду, заметим, что по доказанному выше

$$S = (n-1)\hat{D}_x. \qquad (23)$$

Подставив это выражение в (21), получим

$$|\bar{X}-m_x| < \varepsilon_\alpha \sqrt{\frac{\hat{D}_x}{n}},$$

или

$$\bar{X}-\varepsilon_\alpha \sqrt{\frac{\hat{D}_x}{n}} < m_x < \bar{X}+\varepsilon_\alpha \sqrt{\frac{\hat{D}_x}{n}}. \qquad (24)$$

В таком виде обычно и записывается доверительный интервал для математического ожидания случайной величины.

При больших k T-распределение мало отличается от нормального. Действительно,

$$\ln s_k(t) = \ln c_k - \frac{k+1}{2}\ln\left(1+\frac{t^2}{k}\right) = \ln c_k - \frac{1}{2}t^2 + O\left(\frac{1}{k}\right),$$

где c_k — постоянный множитель в $s_k(t)$. Отсюда видно, что

$$s_k(t) = c_k\left[e^{-\frac{1}{2}t^2} + O\left(\frac{1}{k}\right)\right],$$

и, следовательно, $s_k(t) \to (2\pi)^{-1/2}e^{-t^2/2}$ при $k\to\infty$ ($c_k \to (2\pi)^{-1/2}$, так как $s_k(t)$ — плотность при любом k и сходимость второго множителя к $e^{-t^2/2}$ равномерна на любом конечном интервале). ◀

Таким образом, при больших k (практически при $k > 120$) величину ε_α можно определять, как и при оценке математического ожидания при известной дисперсии, из условия $\Phi(\varepsilon_\alpha) = \alpha/2$, пользуясь таблицей 1 функции Лапласа *).

4.4. Доверительные интервалы для дисперсии. Для нахождения доверительного интервала для дисперсии D_x обычно пользуются первым методом п. 2.4, применяя его к оценке $\sqrt{\hat{D}_x}$ среднего квадратического отклонения $\sigma_x = \sqrt{D_x}$.

▶ Так как величина $U = X^0/\sqrt{D_x}$ имеет нормальное распределение $N(0, 1)$, то согласно результату примера 5.34 величина

$$Z = S/D_x = (U^TU - n\bar{X}^2)/D_x$$

*) При вычислениях на ЭВМ для определения ε_α можно пользоваться программой обращения нормальной функции распределения NDTRI [94].

имеет χ^2-распределение с $k=n-1$ степенями свободы

$$p_k(z)=\frac{1}{2^{k/2}\,\Gamma\left(\frac{k}{2}\right)}\,z^{\frac{k}{2}-1}\,e^{-\frac{1}{2}z}\,1(z).$$

Поэтому, учитывая, что согласно (23) $\hat{D}_x/D_x=S/(kD_x)=Z/k$ и, следовательно,

$$P\left(\frac{1}{1+\varepsilon_\alpha}<\sqrt{\frac{\hat{D}_x}{D_x}}<\frac{1}{(1-\varepsilon_\alpha)_+}\right)=P\left(\frac{k}{(1+\varepsilon_\alpha)^2}<Z<\frac{k}{(1-\varepsilon_\alpha)_+^2}\right),$$

где через $(1-\varepsilon_\alpha)_+$ для краткости обозначена величина $\max(0,\,1-\varepsilon_\alpha)$, получим для ε_α уравнение

$$P\left(\frac{1}{1+\varepsilon_\alpha}<\sqrt{\frac{\hat{D}_x}{D_x}}<\frac{1}{(1-\varepsilon_\alpha)_+}\right)=L_k(\varepsilon_\alpha)=\int\limits_{\frac{k}{(1+\varepsilon_\alpha)^2}}^{\frac{k}{(1-\varepsilon_\alpha)_+^2}} p_k(z)\,dz=\alpha. \quad (25)$$

Отсюда находим доверительный интервал для среднего квадратического отклонения и дисперсии:

$$\max(0,\,1-\varepsilon_\alpha)\sqrt{\hat{D}_x}<\sqrt{D_x}<(1+\varepsilon_\alpha)\sqrt{\hat{D}_x}. \blacktriangleleft \quad (26)$$

Для определения ε_α, удовлетворяющего уравнению (25), составлена таблица значений ε_α в зависимости от k и α (таблица 4).

► При больших k χ-распределение близко к $N\left(\sqrt{k-1/2},\,1/2\right)$. Чтобы доказать это, достаточно показать, что распределение величины $T=\sqrt{2Z}-\sqrt{2k-1}$ стремится к $N(0,\,1)$ при $k\to\infty$. Пользуясь формулой (5.31), находим плотность величины T:

$$f_k(t)=(t+\sqrt{2k-1})\,p_k\left(\frac{(t+\sqrt{2k-1})^2}{2}\right)=\frac{1}{2^{k/2}\,\Gamma\left(\frac{k}{2}\right)}(t+\sqrt{2k-1})^{k-1}e^{-\frac{1}{4}(t+\sqrt{2k-1})^2}.$$

Отсюда находим

$$\ln f_k(t) = \ln a_k + (k-1)\ln\left(t+\sqrt{2k-1}\right) - \frac{1}{4}\left(t+\sqrt{2k-1}\right)^2 = \ln a_k + \frac{k-1}{2}\ln(2k-1) + (k-1)\ln\left(1+\frac{t}{\sqrt{2k-1}}\right) - \frac{1}{4}\left(t+\sqrt{2k-1}\right)^2 =$$

$$= \ln b_k + \frac{k-1}{\sqrt{2k-1}}t - \frac{k-1}{2k-1}\frac{t^2}{2} - \frac{t^2}{4} - \frac{\sqrt{2k-1}}{2}t + O(k^{-1/2}) = \ln b_k - \frac{1}{2}\left(\frac{1}{2}+\frac{k-1}{2k-1}\right)t^2 + O(k^{-1/2}),$$

где через a_k обозначен для краткости постоянный множитель в выражении плотности $f_k(t)$, а $\ln b_k$ — сумма всех постоянных, полученных в результате произведенных преобразований. Из полученной формулы видно, что $f_k(t) \to (2\pi)^{-1/2}e^{-t^2/2}$ при $k \to \infty$, что и доказывает наше утверждение. Практически можно считать величину $\sqrt{Z}$ распределенной нормально при $k > 30$. В этом случае целесообразно отказаться от симметричного относительно оценки $\sqrt{\hat{D}_x}$ доверительного интервала и взять симметричный интервал для статистики $T = \sqrt{2Z} - \sqrt{2k-1}$. Тогда, определив ε_α из уравнения $P(|T| < \varepsilon_\alpha) = 2\Phi(\varepsilon_\alpha) = \alpha$ и имея в виду, что $|t| < \varepsilon_\alpha$ при

$$\frac{(\sqrt{2k-1}-\varepsilon_\alpha)^2}{2} < Z < \frac{(\sqrt{2k-1}+\varepsilon_\alpha)^2}{2},$$

т. е. при

$$\frac{\sqrt{2k-1}-\varepsilon_\alpha}{\sqrt{2k}} < \sqrt{\frac{\hat{D}_x}{D_x}} < \frac{\sqrt{2k-1}+\varepsilon_\alpha}{\sqrt{2k}},$$

получим доверительный интервал для D_x в форме

$$\frac{\sqrt{2k\hat{D}_x}}{\sqrt{2k-1}+\varepsilon_\alpha} < \sqrt{D_x} < \frac{\sqrt{2k\hat{D}_x}}{\sqrt{2k-1}-\varepsilon_\alpha}. \blacktriangleleft \qquad (27)$$

Оценка, распределение которой стремится к нормальному при неограниченном увеличении числа опытов n, называется *асимптотически нормальной*. Таким образом, мы показали, что оценка $\sqrt{\hat{D}_x}$ среднего квадратического отклонения асимптотически нормальна.

4.5. Доверительные области для математического ожидания и дисперсии. Мы нашли доверительные интервалы для математического ожидания и дисперсии по отдельности. Однако при совместном оценивании математического ожидания и дисперсии нас интересует вероятность того, что оба параметра m_x и D_x будут заключены в соответствующих интервалах. Иными словами, нас интересует доверительная область для двумерного векторного параметра (m_x, D_x). Чтобы найти такую область, воспользуемся вторым методом п. 2.4.

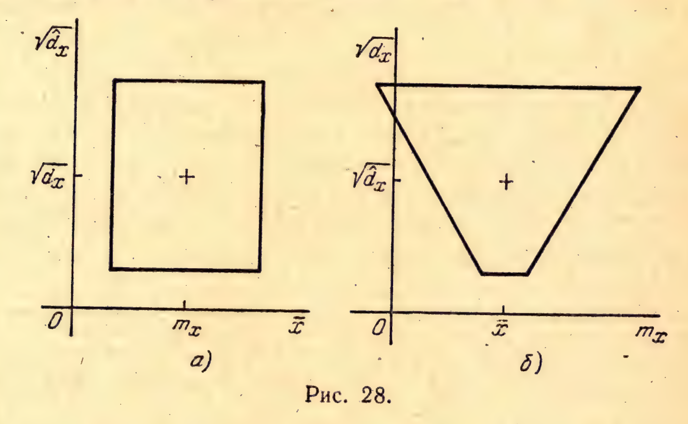


Рис. 28.

► Зададим область $D_\alpha(\theta) = D_\alpha(m_x, D_x)$ на плоскости $(\overline{X}, \sqrt{\hat{D}_x})$ в виде прямоугольника (рис. 28, *а*)

$$D_\alpha(\theta) = \left\{(\overline{X}, \hat{D}_x)\colon \left|\overline{X} - m_x\right| < \varepsilon_\alpha \sqrt{\frac{D_x}{n}},\ \frac{\sqrt{D_x}}{1+\delta_\alpha} < \sqrt{\hat{D}_x} < \frac{\sqrt{D_x}}{(1-\delta_\alpha)_+}\right\},$$

где $(1-\delta_\alpha)_+ = \max(0, 1-\delta_\alpha)$. Тогда область

$$\Delta_\alpha(\hat{\Theta}) = \{(m_x, D_x)\colon\ (\overline{X}, \hat{D}_x) \in D_\alpha(\theta)\}$$

определится формулой

$$\Delta_\alpha(\hat{\Theta}) = \left\{(m_x, D_x)\colon \left|m_x - \overline{X}\right| < \varepsilon_\alpha \sqrt{\frac{D_x}{n}},\right.$$
$$\left.(1-\delta_\alpha)_+ \sqrt{\hat{D}_x} < \sqrt{D_x} < (1+\delta_\alpha)\sqrt{\hat{D}_x}\right\}. \qquad (28)$$

Эта область представляет собой трапецию (рис. 28, *б*).

Чтобы найти вероятность $P((\overline{X}, \hat{D}_x) \in D_\alpha(\theta))$, воспользуемся тем, что при нормальном распределении наблюдаемой случайной величины X статистики $\overline{X}$ и S независимы (пример 5.13). Так как $\overline{X}$ имеет нормальное распределение $N(m_x, D_x/n)$, а $Z = S/D_x = k\hat{D}_x/D_x$ — χ^2-распределение с $k = n-1$ степенями свободы (пример 5.34), то

$$P((\overline{X}, \hat{D}_x) \in D_\alpha(\theta)) = P\left(|\overline{X} - m_x| < \right.$$
$$\left. < \varepsilon_\alpha \sqrt{\frac{D_x}{n}}\right) P\left(\frac{k}{(1+\delta_\alpha)^2} < Z < \frac{k}{(1-\delta_\alpha)_+^2}\right) = 2\Phi(\varepsilon_\alpha) L_k(\delta_\alpha).$$

Определив ε_α и δ_α из уравнений $\Phi(\varepsilon_\alpha) = c\sqrt{\alpha}/2$, $L_k(\delta_\alpha) = \sqrt{\alpha}/c$, где c — произвольно выбранное число, найдем по формуле (28) доверительную область векторного параметра (m_x, D_x), соответствующую коэффициенту доверия α, и можем быть уверены, что в среднем в $100\alpha\%$ случаев точка $(m_x, \sqrt{D_x})$ будет находиться в найденной области $\Delta_\alpha(\hat{\Theta})$. ◀

Пример 6. По данным экспериментальным значениям случайной величины X, полученным в результате 20 опытов (по выборке): 8, 42, 22, —40, 18, 38, 2, —16, 34, 6, 54, 20, 74, 0, 4, —28, 36, 44, 16, 24 — найти оценки и доверительные интервалы для m_x и D_x, соответствующие коэффициенту доверия $\alpha = 0{,}95$.

По формулам (12) и (17) находим оценки m_x, D_x и $\sigma_x = \sqrt{D_x}$:

$$\bar{x} = \frac{358}{20} = 17{,}9, \quad \hat{d}_x = \frac{20\,788 - 20 \cdot 17{,}9^2}{19} \approx 756{,}84,$$
$$\hat{\sigma}_x \approx \sqrt{756{,}84} \approx 27{,}5.$$

По данным $k = n - 1 = 19$ и $\alpha = 0{,}95$ находим по таблицам 3 и 4 $\varepsilon_\alpha = 2{,}09$ для m_x и $\varepsilon_\alpha = 0{,}380$ для D_x. После этого формулы (24) и (26) дают доверительные интервалы для m_x и D_x:

$$5{,}4 < m_x < 30{,}4, \quad 17{,}0 < \sqrt{D_x} < 38{,}0.$$

Для определения доверительной области для векторного параметра (m_x, D_x) находим по таблице функции Лапласа 1 и по таблице 4 $\varepsilon_\alpha = 2{,}23$, $\delta_\alpha = 0{,}464$ из условий $2\Phi(\varepsilon_\alpha) = \sqrt{\alpha} = 0{,}974$, $L_k(\delta_\alpha) = \sqrt{\alpha} = 0{,}974$. После этого по формуле (28) находим доверительную область

$$17{,}9 - 0{,}499\sqrt{D_x} < m_x < 17{,}9 + 0{,}499\sqrt{D_x}, \quad 14{,}7 < \sqrt{D_x} < 40{,}3.$$

На рис. 29 изображена эта трапеция. Для сравнения показан прямоугольник, образованный ранее найденными и доверительными интервалами для m_x и D_x (заштрихован на рис. 29).

Для сравнения найдем прямоугольную доверительную область с коэффициентом доверия не меньше $\alpha=0,95$. Согласно п. 2.4 для этого достаточно найти по формулам (24) и (26) доверительные интервалы для m_x и D_x, соответствующие коэффициенту доверия $\alpha_1=1-(1-\alpha)/2=1-0,05/2=0,975$. В результате получим

$$4,1 < m_x < 31,7,$$

$$14,5 < \sqrt{D_x} < 40,5.$$

Эта прямоугольная доверительная область показана на рис. 29 пунктиром.

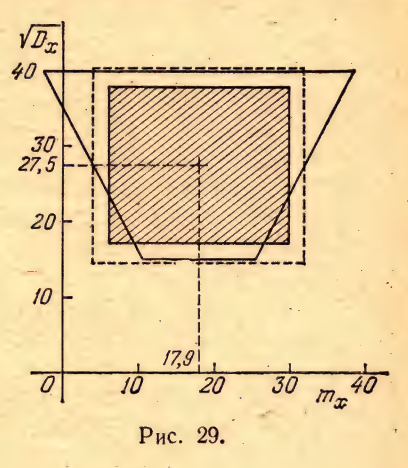


Рис. 29.

4.6. Оценки моментов. На основании результатов, полученных для математических ожиданий, состоятельной несмещенной оценкой момента любого порядка служит среднее арифметическое значение соответствующей степени случайной величины *). Однако точное определение доверительных интервалов для моментов оказывается трудной задачей. Для приближенного определения доверительных интервалов при большом числе опытов обычно пользуются асимптотической нормальностью оценок моментов. На основании центральной предельной теоремы (§ 5.4) распределение оценки момента любого порядка как суммы независимых случайных величин при достаточно большом числе опытов как угодно мало отличается от нормального. Поэтому, зная математическое ожидание и дисперсию оценки момента, можно приближенно определить доверительные интервалы. Математическое ожидание оценки момента равно самому моменту в силу несмещенности оценки. Нахождение дисперсии оценки не представляет никаких принципиальных трудностей, хотя и сопряжено с громоздкими выкладками.

*) Для вычисления оценок математических ожиданий, дисперсий, средних квадратических отклонений, третьего и чётвертого центральных моментов, коэффициентов асимметрии и эксцесса, а также коэффициентов корреляции (п. 5.2) на ЭВМ в [94] служит программа MISR.

§ 5. Оценки математического ожидания и ковариационной матрицы случайного вектора

5.1. Оценки ковариации и коэффициента корреляции. Результаты § 4 позволяют находить оценки и доверительные интервалы для математических ожиданий и дисперсий координат случайного вектора, рассматриваемых по отдельности. Чтобы научиться оценивать ковариационную матрицу случайного вектора, остается рассмотреть оценки ковариаций.

На основании сказанного в п. 1.2.9 естественной оценкой ковариации случайных величин X и Y служит их выборочная ковариация

$$K_{xy}^* = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})(Y_k - \overline{Y}), \tag{29}$$

где $X_1, Y_1, \ldots, X_n, Y_n$ — случайные результаты опытов, а $\overline{X}, \overline{Y}$ — выборочные средние величин X и Y. Совершенно так же, как для оценки D_x^* дисперсии D_x, находим $MK_{xy}^* = (n-1) k_{xy}/n$. Отсюда видно, что K_{xy}^* — смещенная оценка со смещением — k_{xy}/n. Чтобы получить несмещенную оценку $\hat{K}_{xy}$, достаточно умножить K_{xy}^* на $n/(n-1)$. В результате получим

$$\hat{K}_{xy} = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})(Y_k - \overline{Y}). \tag{30}$$

Дисперсия этой оценки выражается через центральные моменты μ_{pq} случайного вектора $\{X, Y\}$:

$$D\hat{K}_{xy} = \frac{n(\mu_{22} - \mu_{11}^2)}{(n-1)^2} - \frac{2\mu_{22} - \mu_{20}\mu_{02} - 3\mu_{11}^2}{(n-1)^2} + \frac{\mu_{22} - \mu_{20}\mu_{02} - 2\mu_{11}^2}{n(n-1)^2}. \tag{31}$$

Эта формула выводится совершенно так же, как (18). В частном случае нормального совместного распределения величин X и Y четвертый момент μ_{22} выражается через элементы ковариационной матрицы $\mu_{22} = \mu_{20}\mu_{02} + 2\mu_{11}^2 = D_x D_y + 2k_{xy}^2$ (п. 4.5.3), и (31) принимает вид

$$D\hat{K}_{xy} = \frac{D_x D_y + k_{xy}^2}{n-1}. \tag{32}$$

На основании (17) и (30) оценку коэффициента корреляции k_{xy} величин X и Y определяют формулой

$$\hat{R}_{xy}^2 = \frac{\hat{K}_{xy}^2}{\hat{D}_x \hat{D}_y} = \frac{\left[\sum_{k=1}^{n} (X_k - \overline{X})(Y_k - \overline{Y})\right]^2}{\sum_{k=1}^{n} (X_k - \overline{X})^2 \sum_{k=1}^{n} (Y_k - \overline{Y})^2}. \qquad (33)$$

5.2. Оценки математического ожидания и ковариационной матрицы. Полученные результаты дают оценки математического ожидания и ковариационной матрицы любого случайного вектора.

▶ Пусть X — m-мерный случайный вектор, $X_1, \ldots, X_n$ — его экспериментальные значения; $X_{k1}, \ldots, X_{km}$ — координаты вектора X_k $(k = 1, \ldots, n)$, $\overline{X}$ — выборочное среднее вектора X:

$$\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k. \qquad (34)$$

Введем матрицу U, строками которой служат векторы $X_1, \ldots, X_n$:

$$U = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1m} \\ \cdot & \cdot & \cdot & \cdot \\ X_{n1} & X_{n2} & & X_{nm} \end{bmatrix}. \qquad (35)$$

Тогда в силу равенства

$$\sum_{k=1}^{n} (X_{kp} - \overline{X}_p)(X_{kq} - \overline{X}_q) = \sum_{k=1}^{n} X_{kp} X_{kq} - n \overline{X}_p \overline{X}_q =$$

$$= \sum_{k=1}^{n} (X_{kp} - m_p)(X_{kq} - m_q) - n(\overline{X}_p - m_p)(\overline{X}_q - m_q), \qquad (36)$$

которое выводится совершенно так же, как аналогичное равенство в п. 4.2, определим оценку ковариационной матрицы K_x формулой

$$\hat{K}_x = \frac{S}{n-1} = \frac{U^T U - n \overline{X}\,\overline{X}^T}{n-1}. \qquad (37)$$

Таким образом, оценкой математического ожидания m_x случайного вектора X может служить его выборочное среднее $\overline{X}$, а несмещенная оценка его ковариационной матрицы K_x определяется формулой (37). ◀

5.3. Доверительные области для математического ожидания. Осталось научиться определять доверительные области для математического ожидания и ковариационной матрицы случайного вектора.

▶ Найдем сначала доверительную область для вектора m_x. Для этого применим третий метод п. 2.4. Согласно результатам примеров 5.15 и 5.25 статистика

$$F=\frac{n(n-m)}{m}(\overline{X}^T-m_x^T)S^{-1}(\overline{X}-m_x)$$

имеет F-распределение

$$f_{m,n-m}(f)=$$

$$=\left(\frac{m}{n-m}\right)^{\frac{m}{2}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-m}{2}\right)}f^{\frac{m}{2}-1}\left(1+\frac{m}{n-m}f\right)^{-\frac{n}{2}}.$$

Эта плотность не зависит от неизвестных параметров m_x и K_x. Поэтому функция

$$\varphi(\overline{X},S,m_x)=F=\frac{n(n-m)}{m}(\overline{X}^T-m_x^T)S^{-1}(\overline{X}-m_x)$$

удовлетворяет всем требованиям п. 2.4, и доверительная область для m_x определяется неравенством

$$(\overline{X}^T-m_x^T)S^{-1}(\overline{X}-m_x)<\frac{m}{n(n-m)}f_\alpha, \tag{38}$$

где f_α — *верхняя* $100(1-\alpha)$-*процентная точка* F-*распределения* $f_{m,n-m}$, определяемая уравнением

$$P(F<f_\alpha)=F_{m,n-m}(f_\alpha)=\int_0^{f_\alpha}f_{m,n-m}(f)\,df=\alpha. \tag{39}$$

Для определения f_α по данным α, $l=m$ и $k=n-m$ составлены таблицы [117] (таблица 6).

Уравнение

$$(\overline{X}^T-\xi^T)S^{-1}(\overline{X}-\xi)=\frac{m}{n(n-m)}f_\alpha$$

в координатах ξ определяет m-мерный эллипсоид (эллипс при $m=2$) с центром в случайной точке $\overline{X}$, случайные размеры и направления главных осей которого определяются матрицей S и числом f_α. Доверительная область (38) представляет собой множество внутренних точек этого случайного эллипсоида. Из (39) следует, что этот

эллипсоид накрывает неизвестную точку m_x с вероятностью α. ◀

5.4. Распределение выборочного коэффициента корреляции. Перейдем к определению доверительных интервалов для коэффициентов корреляции. Из результатов примера 5.18 следует, что в случае нормального распределения вектора X случайная матрица S имеет распределение Уишарта, определяемое плотностью

$$w_{nm}(s) = \frac{c_{nm}}{(2^m |K_x|)^{\frac{n-1}{2}}} |s|^{\frac{n-m-2}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}(K_x^{-1}s)\right\}$$

в области положительно определенных матриц s и $w_{nm}(s) = 0$ вне этой области. Для определения доверительного интервала для коэффициента корреляции достаточно рассмотреть случай двумерного вектора X, $m = 2$. В этом случае совместная плотность элементов S_{11}, S_{12}, S_{22} матрицы S определяется формулой

$$w_{n2}(s_{11}, s_{12}, s_{22}) = \\ = \frac{(s_{11}s_{22} - s_{12}^2)^{\frac{n-4}{2}}}{4\pi(n-3)!\,\Delta^{\frac{n-1}{2}}} \exp\left\{-\frac{1}{2\Delta}(k_{22}s_{11} - 2k_{12}s_{12} + k_{11}s_{22})\right\}, \quad (40)$$

где для краткости положено $\Delta = k_{11}k_{22} - k_{12}^2$. Области симметричных положительно определенных матриц в трехмерном пространстве с прямоугольной системой координат (s_{11}, s_{12}, s_{22}) соответствует область, ограниченная половиной конуса, определяемая неравенствами $s_{11}, s_{22} \geqslant 0$, $s_{12}^2 \leqslant s_{11}s_{22}$. Вне этого конуса $w_{n2}(s_{11}, s_{12}, s_{22}) = 0$.

▶ Чтобы найти распределение оценки (33) коэффициента корреляции, заметим, что $\hat{R}_{12} = S_{12}/\sqrt{S_{11}S_{22}}$, вследствие чего плотность $f_n(\rho)$ оценки $\hat{R}_{12}$ можно определить по формуле (5.32):

$$f_n(\rho) = \int_0^\infty \int_0^\infty w_{n2}(s_{11}, \rho\sqrt{s_{11}s_{22}}, s_{22})\, ds_{11}\, ds_{22}.$$

Подставив сюда выражение (40) и вводя новые переменные интегрирования $x = s_{11}/k_{11}$, $y = s_{22}/k_{22}$, получим после

несложных преобразований

$$f_n(\rho) = \frac{(1-\rho^2)^{\frac{n-4}{2}}}{4\pi(n-3)!\,(1-r^2)^{\frac{n-1}{2}}} \times$$

$$\times \int_0^\infty \int_0^\infty (xy)^{\frac{n-3}{2}} \exp\left\{-\frac{x+y-2r\rho\sqrt{xy}}{2(1-r^2)}\right\} dx\,dy.$$

Замена переменных $y = x\,\mathrm{tg}^2\,\varphi$ дает

$$f_n(\rho) = \frac{(1-\rho^2)^{\frac{n-4}{2}}}{2\pi(n-3)!\,(1-r^2)^{\frac{n-1}{2}}} \times$$

$$\times \int_0^{\pi/2} \frac{\mathrm{tg}^{n-2}\,\varphi}{\cos^2\varphi}\,d\varphi \int_0^\infty x^{n-2} \exp\left\{-\frac{1-r\rho\sin 2\varphi}{2(1-r^2)\cos^2\varphi}\,x\right\} dx.$$

Приняв за новую переменную интегрирования t во внутреннем интеграле модуль аргумента показательной функции вместо x, найдем

$$\int_0^\infty x^{n-2} \exp\left\{-\frac{1-r\rho\sin 2\varphi}{2(1-r^2)\cos^2\varphi}\,x\right\} dx =$$

$$= \frac{2^{n-1}(1-r^2)^{n-1}\cos^{2n-2}\varphi}{(1-r\rho\sin 2\varphi)^{n-1}} \int_0^\infty t^{n-2} e^{-t}\,dt =$$

$$= \frac{2^{n-1}(1-r^2)^{n-1}\cos^{2n-2}\varphi}{(1-r\rho\sin 2\varphi)^{n-1}}\,(n-2)!.$$

Подставив это выражение в предыдущую формулу, получим

$$f_n(\rho) = \frac{n-2}{\pi}(1-r^2)^{\frac{n-1}{2}}(1-\rho^2)^{\frac{n-4}{2}} \int_0^{\pi/2} \frac{\sin^{n-2} 2\varphi\,d\varphi}{(1-r\rho\sin 2\varphi)^{n-1}} =$$

$$= \frac{n-2}{\pi}(1-r^2)^{\frac{n-1}{2}}(1-\rho^2)^{\frac{n-4}{2}} \cdot \frac{1}{2} \int_0^{\pi} \frac{\sin^{n-2}\theta\,d\theta}{(1-r\rho\sin\theta)^{n-1}}.$$

Наконец, принимая во внимание, что $\sin\theta = \sin(\pi-\theta)$, приведем полученную формулу к виду

$$f_n(\rho) = \frac{n-2}{\pi}(1-r^2)^{\frac{n-1}{2}}(1-\rho^2)^{\frac{n-4}{2}} \int_0^{\pi/2} \frac{\sin^{n-2}\theta\,d\theta}{(1-r\rho\sin\theta)^{n-1}}. \blacktriangleleft \quad (41)$$

Эта формула, определяющая плотность выборочного коэффициента корреляции, получена Фишером [110]. Интеграл в (41) вычисляется элементарно (например, подстановкой $u = \operatorname{tg}(\theta/2)$). Однако при этом получится очень громоздкое выражение. Чтобы получить компактную формулу, выведем, следуя Фишеру, рекуррентную формулу для интегралов

$$I_n(z) = \int_0^{\pi/2} \frac{\sin^{n-2}\theta\, d\theta}{(1 - z\sin\theta)^{n-1}}.$$

► Дифференцируя эту формулу по z, получаем

$$I'_n(z) = (n-1) \int_0^{\pi/2} \frac{\sin^{n-1}\theta\, d\theta}{(1 - z\sin\theta)^{n}} = (n-1)\, I_{n+1}(z).$$

Отсюда следует, что $I_n(z) = I'_{n-1}(z)/(n-2)$. Это и есть искомая рекуррентная формула. Применяя ее, находим

$$I_n(z) = I_2^{(n-2)}(z)/(n-2)\,!.$$

Таким образом, задача свелась к вычислению интеграла

$$I_2(z) = \int_0^{\pi/2} \frac{d\theta}{1 - z\sin\theta}.$$

Подстановкой $u = \operatorname{tg}(\theta/2)$ и соответственно $\sin\theta = 2u/(1+u^2)$, $d\theta = 2du/(1+u^2)$ после элементарных тригонометрических преобразований получаем

$$I_2(z) = \int_0^{\pi/2} \frac{d\theta}{1 - z\sin\theta} = \frac{\arccos(-z)}{\sqrt{1 - z^2}}.$$

Таким образом,

$$I_n(z) = \frac{1}{(n-2)\,!} \frac{d^{n-2}}{dz^{n-2}} \frac{\arccos(-z)}{\sqrt{1 - z^2}}.$$

Подставив это выражение в (41), получим формулу Фишера

$$f_n(\rho) = \frac{1}{\pi\,(n-3)\,!} (1 - r^2)^{\frac{n-1}{2}} (1 - \rho^2)^{\frac{n-4}{2}} \times \frac{d^{n-2}}{d(r\rho)^{n-2}} \frac{\arccos(-r\rho)}{\sqrt{1 - r^2\rho^2}}. \quad ◄ \qquad (42)$$

Ясно, что все написанные формулы для $f_n(\rho)$ определяют $f_n(\rho)$ только в интервале $[-1, 1]$. Вне этого интервала $f_n(\rho)=0$.

5.5. Доверительные интервалы для коэффициента корреляции. Для определения доверительного интервала для коэффициента корреляции r обычно пользуются вторым методом п. 2.4, совершенно так же, как это делается для неизвестной вероятности (п. 3.2).

Определив при каждом значении $r \in (-1, 1)$ интервал $D_\alpha(r)=(a_\alpha(r), b_\alpha(r))$ из условия

$$P(\hat{R} < a_\alpha(r)) = P(\hat{R} > b_\alpha(r)) = \frac{1-\alpha}{2}, \tag{43}$$

строим на плоскости $(r, \hat{r})$ кривые $\hat{r}=a_\alpha(r)$ и $\hat{r}=b_\alpha(r)$ (рис. 30). Эти кривые при любом значении $\hat{r}$ оценки $\hat{R}$ коэффициента корреляции определяют соответствующую реализацию $\Delta_\alpha(\hat{r}) = \{r\colon\ a_\alpha(r) < \hat{r} < b_\alpha(r)\}$ доверительного интервала $\Delta_\alpha(\hat{R})$ для r, соответствующего коэффициенту доверия α (рис. 30). Практически кривые $\hat{r}=a_\alpha(r)$ и $\hat{r}=b_\alpha(r)$ можно построить, пользуясь таблицами функции распределения величины $\hat{R}$

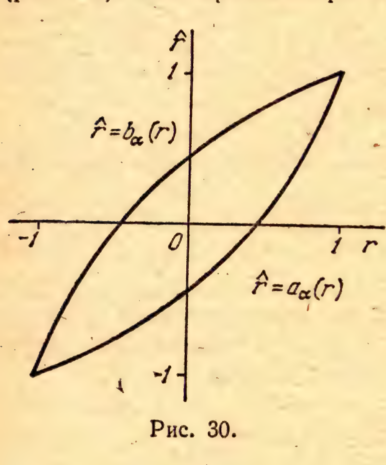


Рис. 30.

$$F_n(\hat{r}) = \int_{-1}^{\hat{r}} f_n(\rho)\, d\rho,$$

которые составлены Дэвид [26, 27].

Для приближенного определения доверительного интервала для r можно воспользоваться тем, что, как заметил Фишер, распределение случайной величины

$$Z = \frac{1}{2} \ln \frac{1+\hat{R}}{1-\hat{R}} \tag{44}$$

даже при небольших значениях n близко к нормальному распределению

$$N\left(\frac{1}{2} \ln \frac{1+r}{1-r} + \frac{r}{2(n-1)},\ \frac{1}{n-3}\right).$$

Дэвид исследовала точность приближения распределения величины Z к нормальному и рекомендует пользоваться приближенным нормальным распределением для Z при $n \geqslant 25$, отбрасывая при этом слагаемое $r/2(n-1)$ в выражении математического ожидания величины Z. Тогда, определив ε_α из условия $P(|Z-m_z| < \varepsilon_\alpha/\sqrt{n-3}) = 2\Phi(\varepsilon_\alpha) = \alpha$, получим приближенный доверительный интервал для коэффициента корреляции r в виде

$$\frac{1}{2}\ln\frac{1+\hat{R}}{1-\hat{R}} - \frac{\varepsilon_\alpha}{\sqrt{n-3}} < \frac{1}{2}\ln\frac{1+r}{1-r} < \frac{1}{2}\ln\frac{1+\hat{R}}{1-\hat{R}} + \frac{\varepsilon_\alpha}{\sqrt{n-3}}.$$

Отсюда, имея в виду, что функция

$$y = \frac{1}{2}\ln\frac{1+x}{1-x}$$

монотонно возрастающая и обратная функция определяется формулой

$$x = \frac{e^{2y}-1}{e^{2y}+1} = \operatorname{th} y,$$

преобразуем доверительный интервал для r к виду

$$\operatorname{th}\left(\frac{1}{2}\ln\frac{1+\hat{R}}{1-\hat{R}} - \frac{\varepsilon_\alpha}{\sqrt{n-3}}\right) < r < \operatorname{th}\left(\frac{1}{2}\ln\frac{1+\hat{R}}{1-\hat{R}} + \frac{\varepsilon_\alpha}{\sqrt{n-3}}\right). \tag{45}$$

5.6. Доверительные области для ковариационной матрицы. Для определения доверительных областей для ковариационной матрицы в $m(m+1)/2$-мерном пространстве можно применить второй метод п.2.4.

Пользуясь распределением Уишарта (5.21), выбираем для каждой данной матрицы K_x область $D_\alpha(K_x)$, вероятность попадания в которую оценки $\hat{K}_x = S/(n-1)$ равна α:

$$P(\hat{K}_x \in D_\alpha(K_x)) = \int\limits_{D_\alpha(K_x)} w_{nm}(s)\,ds = \alpha.$$

Тогда множество матриц K_x, для которых данное значение $\hat{K}_x$ принадлежит $D_\alpha(K_x)$, $\Delta_\alpha(\hat{K}_x) = \{K_x : \hat{K}_x \in D_\alpha(K_x)\}$, будет доверительной областью для K_x, соответствующей коэффициенту доверия α. Область $D_\alpha(K_x)$ можно задать в виде прямоугольника (параллелепипеда) в $m(m+1)/2$-мерном пространстве. Стороны этого прямоугольника по осям $\hat{K}_{pp}$ можно выбрать так, как мы выбирали их в п.4.4 для дисперсии.

▶ Чтобы найти совместную доверительную область для вектора m_x и матрицы K_x (в $m(m+3)/2$-мерном пространстве), воспользуемся тем, что при нормальном распределении наблюдаемого вектора X вектор $\bar{X}$ и матрица S независимы, причем X имеет нормальное распределение $N(m_x, K_x/n)$, а S — распределение Уишарта $w_{nm}(s)$. Будем искать область $D_\alpha(m_x, K_x)$ в виде

$$\{(\bar{X}, \hat{K}_x):\ n(\bar{X}^T - m_x^T)K_x^{-1}(\bar{X} - m_x) < \varepsilon_\alpha,$$
$$a_\alpha(K_x) < \hat{K}_x < b_\alpha(K_x)\}\ ^{*)}.$$

Тогда будем иметь

$$P((\bar{X}, \hat{K}_x) \in D_\alpha(m_x, K_x)) = P(n(\bar{X}^T - m_x^T)K_x^{-1}(X - m_x) <$$
$$< \varepsilon_\alpha)\, P((n-1)\,a_\alpha(K_x) < S < (n-1)\,b_\alpha(K_x)).$$

Определив ε_α, $a_\alpha(K_x)$ и $b_\alpha(K_x)$ из уравнений

$$P(n(\bar{X}^T - m_x^T)\ K_x^{-1}(\bar{X} - m_x) < \varepsilon_\alpha) = c\sqrt{\alpha},$$
$$P((n-1)\,a_\alpha(K_x) < S < (n-1)\,b_\alpha(K_x)) = \frac{\sqrt{\alpha}}{c},$$

получим доверительную область для (m_x, K_x) в виде

$$\Delta_\alpha(\bar{X}, \hat{K}_x) = \{(m_x, K_x):\quad n(m_x^T - \bar{X}^T)K_x^{-1}(m_x - \bar{X}) < \varepsilon_\alpha,$$
$$a_\alpha(K_x) < \hat{K}_x < b_\alpha(K_x)\}.\ ◀$$

Пример 7. В результате 20 опытов получена выборка для двумерного случайного вектора $\{X, Y\}$:

(8, 10), (—9, —17), (—2, —3), (2, —1), (—30, —21), (9, —2), (15, 10), (—10, 0), (—2, —10), (1, 18), (11, —13), (—24, —12), (—8, —2), (27, 18), (6, 5), (—8, —9), (—6, 10), (0, —6), (17, 4), (—18, —20)

Найти оценки математических ожиданий, дисперсий, ковариации и коэффициента корреляции величин X, Y, доверительную область для вектора математического ожидания, доверительные интервалы для дисперсий и коэффициента корреляции и совместные доверительные интервалы для всех пяти параметров m_x, m_y, D_x, D_y, r при $\alpha = 0{,}95$.

По формулам (12), (17), (30) и (33) находим оценки m_x, m_y, D_x, D_y, k_{xy} и r_{xy}:

$$\bar{x} = -1{,}05,\ \bar{y} = -2{,}05,\ \hat{d}_x = s_{11}/19 \approx 194,$$
$$\hat{d}_y = s_{22}/19 \approx 136,\ \hat{k}_{xy} = s_{12}/19 \approx 111,$$
$$\hat{r}_{xy} = \frac{s_{12}}{\sqrt{s_{11}\ s_{22}}} \approx 0{,}68,$$

где $s_{11} = 3681$, $s_{12} = 2111$, $s_{22} = 2583$.

*) Неравенства между матрицами, так же как и между векторами, понимаются как соответствующие неравенства для всех элементов матрицы.

По таблице 6 находим для $\alpha=0{,}95$, $l=m=2$, $k=n-m=18$ значение $f_\alpha=3{,}55$. После этого формула (38) определяет реализацию доверительной области для вектора $[m_x \; m_y]^T$ как множество внутренних точек эллипса

$$0{,}5113\,(m_x+1{,}05)^2-0{,}8358\,(m_x+1{,}05)\,(m_y+2{,}05)+ \\ +0{,}7287\,(m_y+2{,}05)^2=19{,}72.$$

На рис. 31 этот эллипс показан вместе со всеми экспериментальными точками.

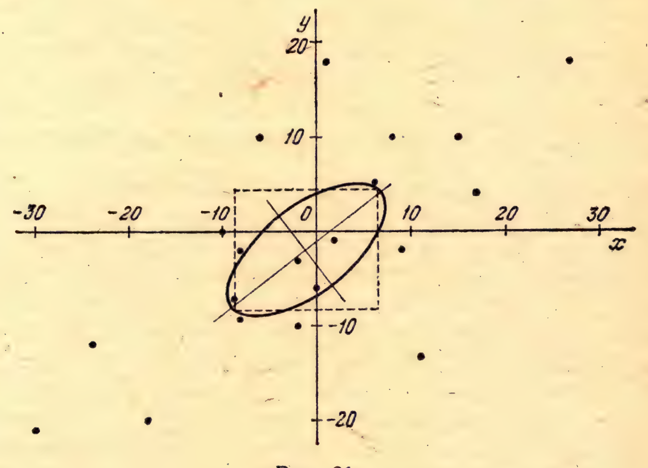


Рис. 31.

Определив по таблице 4 значение $\varepsilon_\alpha=0{,}380$ по данным $\alpha=0{,}95$, $k=n-1=19$, находим по формуле (26) доверительные интервалы для дисперсий координат случайного вектора:

$$8{,}6<\sqrt{D_x}<19{,}2,\quad 7{,}2<\sqrt{D_y}<16{,}1.$$

Наконец, определив по таблице 1 значение $\varepsilon_\alpha=1{,}96$, для которого $2\Phi(\varepsilon_\alpha)=\alpha=0{,}95$, находим по приближенной формуле (45) доверительный интервал для коэффициента корреляции:

$$\operatorname{th}\left(\frac{1}{2}\ln\frac{1{,}68}{0{,}32}-\frac{1{,}96}{\sqrt{17}}\right)<r<\operatorname{th}\left(\frac{1}{2}\ln\frac{1{,}68}{0{,}32}+\frac{1{,}96}{\sqrt{17}}\right),$$

или, выполнив вычисления,

$$0{,}338<r<0{,}862.$$

Совместные доверительные интервалы для m_x, m_y, D_x, D_y и r, соответствующие $\alpha=0{,}95$, согласно концу п.2.4 можно определить как доверительные интервалы для каждого из этих параметров, соответствующие коэффициенту доверия $\alpha_1=1-(1-\alpha)/5=0{,}99$.

В результате получаем

$$-20{,}96 < m_x < 18{,}86, \quad -18{,}72 < m_y < 14{,}62,$$
$$5{,}8 < \sqrt{D_x} < 22{,}0, \quad 4{,}9 < \sqrt{D_y} < 18{,}5,$$
$$0{,}200 < r < 0{,}895.$$

Найдем еще совместные доверительные интервалы для величин m_x и m_y. Для этого следует определить доверительные интервалы для каждой из величин m_x, m_y, соответствующие коэффициенту доверия $\alpha_2 = 1 - (1 - \alpha)\,2 = 0{,}975$. В результате получим

$$-8{,}80 < m_x < 6{,}70, \quad -8{,}47 < m_y < 4{,}37.$$

Для сравнения на рис. 31 показан прямоугольник, образованный совместными доверительными интервалами для m_x и m_y.

§ 6. Проверка гипотез о параметрах распределений

6.1. Задачи проверки гипотез. Задаче построения доверительных областей для параметров распределений родственна задача проверки гипотез об этих параметрах. Из сказанного в п. 1.1 ясно, что никаких точных утверждений о параметрах распределения по результатам опытов делать нельзя. Можно лишь высказывать различные предположения о них — *гипотезы*. Задача проверки гипотез состоит в том, чтобы установить, противоречит принятая гипотеза экспериментальным данным или нет.

Возможны два основных типа гипотез о неизвестных параметрах распределения. Гипотеза первого типа представляет собой предположение, что неизвестный параметр θ (скалярный или векторный) имеет данное значение, скажем θ_0, или принадлежит данному множеству значений. Гипотеза второго типа состоит в том, что неизвестные параметры в двух независимых выборках (или в большем числе выборок) имеют одни и те же значения. Ясно, что гипотеза второго типа, по существу, представляет собой предположение, что две серии опытов, в которых получены выборки, произведены в одних и тех же условиях, так как только при выполнении определенного комплекса условий в каждом опыте вероятности событий и характеристики случайных величин имеют одни и те же значения во всех опытах. Характерной причиной изменения условий опыта часто являются различного рода *дрейфы* (*тренды*) — медленные неконтролируемые изменения различных величин, характеризующих условия опытов. При этом изменение величины считается медленным, если в течение любой одной серии

опытов ее можно считать практически постоянной и только в интервале времени между двумя сериями опытов ее изменение может быть существенным. Задача проверки гипотез второго типа в таких случаях представляет собой задачу обнаружения существенных дрейфов за время между двумя сериями опытов.

Алгоритм, в соответствии с которым экспериментальным данным ставится в соответствие решение принять или отвергнуть гипотезу, называется *правилом решения* или *решающим правилом*.

6.2. Проверка гипотез о значении параметра. Задача проверки гипотез первого типа легко решается с помощью доверительных областей. Если данное значение θ_0 (данное множество значений) принадлежит полученной реализации доверительной области (имеет непустое пересечение с ней), то можно считать, что гипотеза не противоречит экспериментальным данным. В противном случае гипотеза отвергается. При таком решающем правиле вероятность отклонить гипотезу в случае, когда она верна, равна $1-\alpha$.

Пример 8. В условиях примера 5 гипотезу о том, что вероятность события A равна 0,5, следует отвергнуть, так как точка $p=0,5$ не принадлежит доверительному интервалу (0,583; 0,763), соответствующему $\alpha=0,95$. Вероятность отвергнуть правильную гипотезу в данном случае равна 0,05.

Пример 9. В условиях примера 7 гипотезу о некоррелированности случайных величин следует отвергнуть, а гипотезу о том, что коэффициент корреляции заключен в интервале (0; 0,5), можно принять, так как точка $r=0$ не принадлежит доверительному интервалу (0,338; 0,862), а интервал (0; 0,5) пересекается с этим доверительным интервалом.

Другой подход к задаче проверки гипотез первого типа будет дан в § 10.5.

6.3. Проверка гипотез о совпадении значений параметра. Для решения задачи проверки гипотез второго типа часто применяется видоизмененный третий метод п. 2.4. Видоизменение состоит лишь в том, что вместо неизвестного параметра θ в число аргументов функции φ включается его оценка по второй выборке.

Пусть $\hat{\Theta}_1$ и $\hat{\Theta}_2$ — две оценки параметра θ, полученные по двум независимым выборкам. Предположим, что удалось найти скалярную функцию $\varphi(\hat{\Theta}_1, \hat{\Theta}_2, S)$ оценок $\hat{\Theta}_1$, $\hat{\Theta}_2$ и некоторой статистики S, обладающую следующими свойствами:

1) при любых фиксированных значениях $\hat{\theta}_1$ и s величин $\hat{\Theta}_1$ и S неравенство $\varphi(\hat{\theta}_1, \hat{\theta}_2, s) < c$ при возрастании c определяет монотонно возрастающее семейство вложенных одна в другую областей $D(\hat{\theta}_1, s, c) = \{\hat{\theta}_2: \varphi(\hat{\theta}_1, \hat{\theta}_2, s) < c\}$;

2) $\varphi(\hat{\theta}, \hat{\theta}, s) = 0$ при любых $\hat{\theta}$, s и $\varphi(\hat{\theta}_1, \hat{\theta}_2, s) > 0$ при любых $\hat{\theta}_1$, $\hat{\theta}_2$, s, $\hat{\theta}_1 \neq \hat{\theta}_2$ (и, следовательно, точка $\hat{\theta}_2 = \hat{\theta}_1$ принадлежит области $D(\hat{\theta}_1, s, c)$ при любых $c > 0$, $\hat{\theta}_1$ и s);

3) распределение случайной величины $T = \varphi(\hat{\Theta}_1, \hat{\Theta}_2, S)$ в случае справедливости гипотезы не зависит от неизвестного параметра θ.

Тогда, определив δ_α из уравнения

$$P(\varphi(\hat{\Theta}_1, \hat{\Theta}_2, S) < \delta_\alpha) = \alpha,$$

получим следующее *правило решения*, соответствующее коэффициенту доверия α: если $\varphi(\hat{\Theta}_1, \hat{\Theta}_2, S) < \delta_\alpha$, то гипотеза не противоречит экспериментальным данным; если же $\varphi(\hat{\Theta}_1, \hat{\Theta}_2, S) \geqslant \delta_\alpha$, то гипотеза отвергается. При этом вероятность отвергнуть гипотезу в случае, когда она верна, равна $1-\alpha$.

Применим изложенный общий метод для проверки гипотезы, что математическое ожидание случайной величины одно и то же (неизвестное) значение в двух независимых сериях опытов. Рассмотрим две серии независимых наблюдений нормально распределенной величины X (или двух случайных величин X_1 и X_2) с неизвестным математическим ожиданием m_x и ковариационной матрицей K_x. Гипотеза, подлежащая проверке, состоит в том, что m_x предполагается одинаковым в обеих сериях опытов. Матрица K_x при этом считается одной и той же в обеих сериях опытов. Пусть n_1 и n_2 — число опытов в первой и второй сериях, $\bar{X}_1$ и $\bar{X}_2$ — случайные выборочные средние в двух сериях опытов, U_1 и U_2 — матрицы, столбцами которых служат выборочные значения наблюдаемого случайного вектора X в двух сериях опытов,

$$S_k = U_k U_k^T - n_k \bar{X}_k \bar{X}_k^T = U_k^0 U_k^{0T} - n_k \bar{X}_k^0 \bar{X}_k^{0T} \quad (k = 1,\ 2),$$
$$H = S_1 + S_2,$$

m — размерность наблюдаемого вектора X.

Согласно результатам примеров 5.17 и 5.25 случайная величина

$$F=\frac{n_1n_2(n_1+n_2-m-1)}{m(n_1+n_2)}(\bar{X}_1^T-\bar{X}_2^T)H^{-1}(\bar{X}_1-\bar{X}_2)$$

имеет F-распределение $f_{m,\,n_1+n_2-m-1}(f)$. Так как это распределение не зависит от неизвестных m_x и K_x, то функция

$$\varphi(\bar{X}_1,\ \bar{X}_2,\ H)=\frac{n_1n_2(n+n_2-m-1)}{m(n_1+n_2)}(\bar{X}_1^T-\bar{X}_2^T)H^{-1}(\bar{X}_1-\bar{X}_2)$$

удовлетворяет всем необходимым условиям. Поэтому можно принять следующее решающее правило для проверки гипотезы о равенстве математических ожиданий в двух сериях опытов: если

$$(\bar{X}_1^T-\bar{X}_2^T)H^{-1}(\bar{X}_1-\bar{X}_2)<f_\alpha\frac{m(n_1+n_2)}{n_1n_2(n_1+n_2-m-1)}, \qquad (46)$$

где f_α — верхняя $100(1-\alpha)$-процентная точка F-распределения $f_{m,\,n_1+n_2-m-1}$, то гипотеза принимается; в противном случае она отвергается. В первом случае говорят, что расхождение двух средних *незначимо* и может быть вызвано только случайностью результатов опытов. Во втором случае говорят, что расхождение *значимо* и маловероятно, что оно вызвано только случайностью результатов опытов. Значимое расхождение указывает на возможность различия математических ожиданий в двух выборках (вследствие неконтролируемого различия в условиях опытов). Величина f_α практически определяется по таблице 6 по данным α, $l=m$ и $k=n_1+n_2-m-1$.

В частном случае скалярной наблюдаемой величины X $(m=1)$ можно преобразовать полученное решающее правило так, чтобы вместо F-распределения пользоваться T-распределением. На основании результатов примера 5.14 случайная величина

$$T=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{(n_1+n_2)H}}(\bar{X}_1-\bar{X}_2)$$

имеет T-распределение $s_{n_1+n_2-2}(t)$. Так как это распределение не зависит от m_x и $D_x=K_x$, то за функцию $\varphi(\hat{X}_1,\ \hat{X}_2,\ H)$ можно принять статистику T. Тогда получим следующее правило: если

$$|\bar{X}_1-\bar{X}_2|<t_\alpha\sqrt{\frac{(n_1+n_2)H}{n_1n_2(n_1+n_2-2)}}, \qquad (47)$$

где t_α определяется уравнением $P(|T| < t_\alpha) = \alpha$, то гипотеза принимается; в противном случае она отвергается, расхождение между двумя средними признается значимым. Практически t_α определяется по таблице 3 по данным α и $k = n_1 + n_2 - 2$ *). Ясно, что (46) при $m = 1$ совпадает с (47) (пример 5.15).

Для дополнительного подкрепления решения о том, что расхождение двух средних незначимо, можно вычислить вероятность получения расхождения, большего чем наблюденное. В случае скалярной величины X эту вероятность $P(|T| > |t|)$, где t — полученная в результате опытов реализация величины T, можно найти по таблице 3. Значение этой вероятности $P(F > f)$ в общем случае согласно результату примера 5.26 равно значению функции распределения для β-распределения с параметрами $p = (n_1 + n_2 - m - 1)/2$, $q = m/2$ в точке $v = [1 + mf/(n_1 + n_2 - m - 1)]^{-1}$, где f — полученная в результате опытов реализация величины F **).

Пример 10. В условиях примера 6 по второй выборке, содержащей результаты 30 опытов, получено среднее 23,5 и $s_2 = 354$. Проверить гипотезу о равенстве математических ожиданий в двух сериях опытов, приняв $\alpha = 0{,}95$.

В данном случае $n_1 = 20$, $n_2 = 30$, $m = 1$, $\bar{x}_1 = 17{,}9$, $\bar{x}_2 = 23{,}5$, $s_1 = 14\,380$, $s_2 = 23\,354$, $h = s_1 + s_2 = 37\,734$, $t_\alpha = 2{,}013$ и

$$|\bar{x}_1 - \bar{x}_2| = |17{,}9 - 23{,}5| = 5{,}6 < 2{,}013 \sqrt{\frac{50 \cdot 37\,734}{20 \cdot 30 \cdot 48}}.$$

Поэтому гипотезу о совпадении математических ожиданий можно принять. Для подкрепления этого решения найдем значение t величины T, полученное в результате опытов:

$$t = \sqrt{\frac{20 \cdot 30 \cdot 48}{50 \cdot 37734}}(17{,}9 - 23{,}5) \approx -0{,}692.$$

После этого по таблице 3 находим для $k = 48$ $P(|T| > 0{,}692) \approx 0{,}493$. Таким образом, с вероятностью 0,5 можно получить еще большее расхождение средних, чем наблюденное.

6.4. Исключение аномальных наблюдений. Изложенный метод проверки гипотезы о совпадении математических ожиданий в двух выборках можно, в частности, применить для проверки гипотезы об аномальности на-

---

*) Для вычисления величины T для проверки гипотез о равенстве m_x данному числу и о равенстве двух математических ожиданий служит программа TTEST [94].

**) Для β-распределения функцию распределения можно вычислить на ЭВМ по программе BDTR [94].

блюдения, давшего в результате изолированную точку, удаленную от группы остальных экспериментальных точек. При таком «выскоке» экспериментальной точки из группы остальных точек естественно возникает мысль, что этот «выскок» произошел из-за неконтролируемого нарушения условий опыта и поэтому «выскочившую» точку следует отбросить, признать аномальной. Ясно, что для принятия такого решения нужны веские основания. Так возникает задача проверки гипотезы об аномальности наблюдения.

Предположим, что в результате $n+1$ опытов одно значение случайной величины оказывается на значительном удалении от группы n остальных значений. Без потери общности можно считать, что отклонение произошло в $(n+1)$-м опыте. Для проверки гипотезы об аномальности значения x_{n+1} достаточно рассматривать его как вторую выборку, состоящую из одной экспериментальной точки x_{n+1}. Тогда можно будет применить изложенный метод. В данном случае $n_1 = n$, $n_2 = 1$, $\bar{X}_2 = X_{n+1}$, $S_2 = 0$ и (46) дает следующее решающее правило: значение X_{n+1} наблюдаемой величины X признается аномальным и отбрасывается, если

$$(\bar{X}_1^T - X_{n+1}^T) S_1^{-1} (\bar{X}_1 - X_{n+1}) \geqslant f_\alpha \frac{m\,(n+1)}{n\,(n-m)},$$

где f_α — верхняя $100(1-\alpha)$-процентная точка F-распределения $f_{m,\,n-m}(f)$; в противном случае оно учитывается.

Точно так же в случае скалярной величины X (47) дает правило: значение X_{n+1} признается аномальным и отбрасывается, если

$$|\bar{X}_1 - X_{n+1}| \geqslant t_\alpha \sqrt{\frac{(n+1)\,S_1}{n\,(n-1)}},$$

где t_α определяется по таблице 3 по данным α и $k = n-1$; в противном случае X_{n+1} учитывается.

Пример 11. В результате 21 опыта получены 20 экспериментальных точек примера 7 и еще одна точка (50, 40), удаленная от группы остальных. Проверить гипотезу об аномальности этой точки.

По таблице 6 для $\alpha = 0{,}95$, $l = m = 2$, $k = n - m = 18$ находим $f_\alpha = 3{,}55$. Так как

$$(\bar{x}_1^T - \bar{x}_2^T) s_1^{-1} (\bar{x}_1 - \bar{x}_2) = [0{,}5113\,(50+1{,}05)^2 - 0{,}8358\,(50+1{,}05)\,(40+2{,}05) + 0{,}7287\,(40+2{,}05)^2] \cdot 10^{-3} \approx 0{,}827 > 3{,}55 \cdot \frac{2 \cdot 21}{20 \cdot 18} \approx 0{,}414,$$

то точку $x_{21} = [50\ 40]^T$ можно признать аномальной и исключить.

Глава 7

ТЕОРИЯ ОЦЕНОК

§ 1. Общие свойства оценок

1.1. Некоторые соотношения. Для каждой статистической характеристики можно найти много различных оценок. Мы это видели на примере оценок дисперсий и ковариаций. Другие оценки для дисперсии нормально распределенной случайной величины можно получить, выразив дисперсию $D_x = \mu_2$ из формулы (3.89) через любой четный центральный момент μ_{2k} и заменив в полученном равенстве момент μ_{2k} его оценкой. За оценку математического ожидания случайной величины с симметричным распределением можно принять ее выборочную *медиану*, т. е. среднее $X_{\nu+1}$ из экспериментальных значений величины X, пронумерованных в порядке возрастания, при нечетном числе опытов $n = 2\nu + 1$ и полусумму двух средних $(X_\nu + X_{\nu+1})/2$ или величину $\lambda X_\nu + (1-\lambda) X_{\nu+1}$, $\lambda \in (0, 1)$, при четном числе опытов $n = 2\nu$. Естественно стремиться находить в известном смысле наилучшие оценки с минимальной дисперсией или смещенные оценки с минимальным средним квадратом ошибки (§ 6.2). Для этого необходимо изучить общие свойства оценок и установить некоторые общие методы их нахождения.

Как и прежде, будем рассматривать результаты опытов (не обязательно независимых) $X_1, \ldots, X_n$ как координаты n-мерного случайного вектора (строки случайной $n \times m$-матрицы в случае m-мерного наблюдаемого вектора X) U и обозначим через $g(u \mid \theta)$ плотность величины U, зависящую от неизвестного параметра θ. Любая оценка параметра θ согласно определению п. 6.2.1 представляет собой функцию результатов опытов U, не зависящую от θ: $\hat{\Theta} = \varphi(U)$. Математическое ожидание этой

оценки в общем случае зависит от параметра θ:

$$M\hat{\Theta} = M\varphi(U) = \int_{-\infty}^{\infty} \varphi(u) g(u \mid \theta) du = m_\varphi(\theta). \qquad (1)$$

Для несмещенной оценки $\hat{\Theta}$ по определению $m_\varphi(\theta) = \theta$.

▶ Будем считать в общем случае параметр θ r-мерным вектором и соответственно через $\partial/\partial\theta$ обозначать оператор градиента: $\partial/\partial\theta = [\partial/\partial\theta_1 \ldots \partial/\partial\theta_r]^T$. Тогда будем иметь $\partial^T/\partial\theta = [\partial/\partial\theta_1 \ldots \partial/\partial\theta_r]$ и, дифференцируя формулу (1) по θ, получим

$$m'_\varphi(\theta) = \int_{-\infty}^{\infty} \varphi(u) \frac{\partial^T g(u \mid \theta)}{\partial\theta} du =$$

$$= \int_{-\infty}^{\infty} \varphi(u) \frac{\partial^T \ln g(u \mid \theta)}{\partial\theta} g(u \mid \theta) du = M\left[\hat{\Theta}\left[\frac{\partial^T \ln g(U \mid \theta)}{\partial\theta}\right]\right]. \qquad (2)$$

Здесь, как и в § 5.1, $m'_\varphi(\theta)$ представляет собой квадратную матрицу с элементами $m'_{pq}(\theta) = \partial m_p(\theta)/\partial\theta_q$, где $m_1(\theta), \ldots, m_r(\theta)$ — координаты вектора $m_\varphi(\theta)$. При этом производные функции $\ln g(u \mid \theta)$ по $\theta_1, \ldots, \theta_r$ могут существовать в обычном смысле или как обобщенные функции, содержащие линейные комбинации δ-функций. Точно так же, дифференцируя формулу

$$\int_{-\infty}^{\infty} g(u \mid \theta) du = 1,$$

находим

$$\int_{-\infty}^{\infty} \frac{\partial g(u \mid \theta)}{\partial\theta} du = \int_{-\infty}^{\infty} \frac{\partial \ln g(u \mid \theta)}{\partial\theta} g(u \mid \theta) du = 0, \qquad (3)$$

или

$$M \frac{\partial \ln g(U \mid \theta)}{\partial\theta} = 0. \qquad (4)$$

Отсюда и из (2) следует формула

$$M(\hat{\Theta} - \theta) Z^T = M\left[(\hat{\Theta} - \theta) \frac{\partial^T \ln g(U \mid \theta)}{\partial\theta}\right] = m'_\varphi(\theta). \qquad (5)$$

Таким образом, *случайный вектор*

$$Z = \frac{\partial \ln g(U \mid \theta)}{\partial\theta}$$

имеет нулевое математическое ожидание и его ковариация с ошибкой $\hat{\Theta}-\theta$ *оценки* $\hat{\Theta}=\varphi(U)$ *равна матрице производных* $m'_\varphi(\theta)$ *вектора* $M\hat{\Theta}=m_\varphi(\theta)$ *по* $\theta_1, \ldots, \theta_r$. ◀

1.2. Нижняя грань рассеивания оценки скалярного параметра. Из равенства (5) можно получить нижнюю грань рассеивания оценки $\hat{\Theta}$.

▶ Рассмотрим сначала случай скалярного параметра θ. В этом случае все величины в (5) скалярные и к математическому ожиданию в левой части можно применить неравенство (3.45) для моментов второго порядка скалярных случайных величин. В результате получим

$$[m'_\varphi(\theta)]^2 \leqslant M(\hat{\Theta}-\theta)^2 \cdot MZ^2,$$

или

$$[m'_\varphi(\theta)]^2 \leqslant M(\hat{\Theta}-\theta)^2 \cdot DZ. \tag{6}$$

Отсюда в случае, когда $DZ < \infty$, получаем неравенство

$$M(\hat{\Theta}-\theta)^2 \geqslant \frac{[m'_\varphi(\theta)]^2}{DZ} = \frac{[m'_\varphi(\theta)]^2}{D\partial \ln g/\partial\theta}, \tag{7}$$

где для краткости опущены аргументы U и θ функции g. Это неравенство впервые получено Фишером [111] (см. также [28, 45, 84, 114]).

В частном случае несмещенной оценки $\hat{\Theta}$ $m'_\varphi(\theta)=1$, $M(\hat{\Theta}-\theta)^2=D\hat{\Theta}$ и неравенство (7) принимает вид

$$D\hat{\Theta} \geqslant \frac{1}{DZ} = \frac{1}{D\partial \ln g/\partial\theta}. \tag{8}$$ ◀

Это неравенство определяет нижнюю грань (не обязательно точную) дисперсии несмещенной оценки. Никакая несмещенная оценка не может иметь меньшую дисперсию, чем правая часть (8).

Если $DZ=\infty$, то неравенства (6) и (7) становятся тривиальными и не определяют нижнюю грань среднего квадрата ошибки. Это имеет место, например, в случае разрывной функции $g(u|\theta)$ с зависящими от θ точками разрыва. В этом случае производная $\partial \ln g/\partial\theta$ содержит δ-функции, а $(\partial \ln g/\partial\theta)^2$ не существует (квадрат δ-функции не имеет смысла).

1.3. Эффективная оценка скалярного параметра. Оценка $\hat{\Theta}$, для которой в (7) имеет место знак равенства, называется *эффективной*.

▶ По доказанному в § 3.3 знак равенства в (6), а следовательно, и в (7) получается тогда и только тогда, когда случайные величины $\hat{\Theta}-\theta$ и $Z=\partial \ln g/\partial\theta$ связаны линейной зависимостью:

$$Z=\frac{\partial \ln g(U\,|\,\theta)}{\partial\theta}=c(\hat{\Theta}-\theta)=c[\varphi(U)-\theta], \qquad (9)$$

где коэффициент пропорциональности c может зависеть от θ, но не зависит от U. Отсюда и из (4) следует, что математическое ожидание эффективной оценки $\hat{\Theta}$ всегда равно θ, $M\hat{\Theta}=\theta$. Таким образом, *любая эффективная оценка является несмещенной*. ◀

Коэффициент c в (9) легко определяется. Имея в виду, что для эффективной оценки $D\hat{\Theta}=1/DZ$, из (9) находим

$$DZ=c^2 D\hat{\Theta}=c^2/DZ,$$

откуда следует $c=DZ=D\partial \ln g/\partial\theta$.

Ясно, что эффективная оценка существует тогда и только тогда, когда функция $\partial \ln g(u\,|\,\theta)/\partial\theta$ представима в виде (9). В этом случае правая часть неравенства (8) представляет собой точную нижнюю грань дисперсии для несмещенных оценок:

$$\frac{1}{DZ}=\frac{1}{D\partial \ln g/\partial\theta}=\inf_{M\hat{\Theta}=\theta} D\hat{\Theta}. \qquad (10)$$

Если функция $\partial \ln g/\partial\theta$ не может быть представлена в виде (9), то эффективной оценки не существует. В этом случае правая часть в (8) не является точной нижней гранью дисперсии несмещенной оценки.

Заметим, что из несмещенности любой эффективной оценки следует, что ни для какой смещенной оценки в (7) не может быть знака равенства. Тем не менее, во всех случаях, когда существует эффективная оценка, существует смещенная оценка более точная, чем эффективная, т. е. с меньшим средним квадратом ошибки. Однако несмещенными оценками обычно не пользуются, чтобы избежать систематических ошибок при небольшом числе опытов n. При большом же n заметного выигрыша в точности по сравнению с эффективной оценкой не получается. Поэтому эффективными оценками пользуются всегда, когда они существуют.

Для любой несмещенной оценки $\hat{\Theta}=\varphi(U)$ отношение правой части неравенства (8) к левой называется *эффективностью* этой оценки и обозначается $e(\varphi)$:

$$e(\varphi)=\frac{1}{D\hat{\Theta}\cdot DZ}=\frac{1}{D\hat{\Theta}\cdot D\partial \ln g/\partial\theta}. \tag{11}$$

Эффективность любой эффективной оценки равна 1. Эффективность любой другой несмещенной оценки представляет собой положительное число, меньшее единицы.

Из (9) ясно, что если эффективная оценка существует, то она единственная, так как формула (9) не может быть справедлива для двух различных функций φ.

Оценка, эффективность которой стремится к единице при неограниченном возрастании числа опытов, называется *асимптотически эффективной*.

1.4. Нижняя грань рассеивания оценки векторного параметра. Перейдем к случаю векторного параметра θ. Перепишем (7) в виде

$$\frac{\xi^2}{\gamma_\varphi}\leqslant\frac{DZ}{m_\varphi'^2}\xi^2, \tag{12}$$

где $\gamma_\varphi=M(\hat{\Theta}-\theta)^2$ — второй начальный момент ошибки оценки $\hat{\Theta}=\varphi(U)$, а ξ — вспомогательная переменная. Таким образом, квадратичная форма одной скалярной переменной ξ в левой части (12) ни при каких значениях ξ не может быть больше квадратичной формы в правой части. Это наводит на мысль, что, построив аналогичные квадратичные формы для случая векторного параметра θ, мы сможем распространить этот факт на многомерные оценки.

Пусть $\hat{\Theta}=\varphi(U)$ — оценка r-мерного векторного параметра θ. Момент второго порядка ошибки этой оценки определяется формулой

$$\Gamma_\varphi=M(\hat{\Theta}-\theta)(\hat{\Theta}^T-\theta^T).$$

Естественно предполагать, что между координатами вектора $\hat{\Theta}-\theta$ не существует линейных зависимостей. В этом случае матрица Γ_φ обратима. Естественно также считать, что между координатами вектора $m_\varphi(\theta)$ не существует зависимостей вида $\varphi(m_1(\theta), \ldots, m_r(\theta))=0$. В таком случае матрица $m_\varphi'(\theta)$ обратима. При этих условиях обобщение неравенства (12) на случай векторного параметра θ имеет вид

$$\xi^T\Gamma_\varphi^{-1}\xi\leqslant\xi^T\mu_\varphi^T K_z\mu_\varphi\xi, \tag{13}$$

где через μ_φ для краткости обозначена матрица, обратная m'_φ, $\mu_\varphi = (m'_\varphi)^{-1}$, а ξ — вспомогательная векторная переменная (матрица-столбец).

▶ Чтобы доказать (13), вычислим момент второго порядка случайного вектора

$$V = \Gamma_\varphi^{-1}(\hat{\Theta} - \theta) - \mu_\varphi^T Z.$$

Имеем

$$\Gamma_v = MVV^T = \Gamma_\varphi^{-1} M(\hat{\Theta} - \theta)(\hat{\Theta}^T - \theta^T)\Gamma_\varphi^{-1} - \\ - \Gamma_\varphi^{-1} M(\hat{\Theta} - \theta) Z^T \mu_\varphi - \mu_\varphi^T MZ(\hat{\Theta}^T - \theta^T)\Gamma_\varphi^{-1} + \mu_\varphi^T MZZ^T \mu_\varphi.$$

Отсюда, учитывая, что

$$M(\hat{\Theta} - \theta)(\hat{\Theta}^T - \theta^T) = \Gamma_\varphi, \quad MZZ^T = K_z,$$
$$M(\hat{\Theta} - \theta) Z^T = m'_\varphi = \mu_\varphi^{-1}, \quad MZ(\hat{\Theta}^T - \theta^T) = m'^T_\varphi = (\mu_\varphi^T)^{-1},$$

получаем

$$\Gamma_v = \mu_\varphi^T K_z \mu_\varphi - \Gamma_\varphi^{-1}.$$

Но любой момент второго порядка представляет собой неотрицательно определенную матрицу (§ 3.3), вследствие чего $\xi^T \Gamma_v \xi \geqslant 0$ при всех ξ, откуда и следует (13). В частном случае несмещенной оценки $\hat{\Theta}$ $m_\varphi(\theta) = \theta$, $m'_\varphi(\theta) = I$, $\Gamma_\varphi = K_\varphi$ и неравенство (13) принимает вид

$$\xi^T K_\varphi^{-1} \xi \leqslant \xi^T K_z \xi. \; ◀ \tag{14}$$

Геометрически (14) означает, что эллипсоид $\xi^T K_z \xi = c$ при любом c расположен целиком внутри эллипсоида $\xi^T K_\varphi^{-1} \xi = c$ для любой оценки $\hat{\Theta} = \varphi(U)$. Следовательно, эллипсоид рассеивания $\xi^T K_\varphi^{-1} \xi = 1$ любой несмещенной оценки $\hat{\Theta} = \varphi(U)$ ограничен изнутри фиксированным эллипсоидом $\xi^T K_z \xi = 1$.

Ковариационная матрица

$$K_z = M \frac{\partial \ln g}{\partial \theta} \frac{\partial^T \ln g}{\partial \theta}$$

случайного вектора $Z = \partial \ln g / \partial \theta$ называется *информационной матрицей Фишера*.

1.5. Эффективная оценка векторного параметра. Если в (13) имеет место знак равенства, то оценка $\hat{\Theta} = \varphi(U)$ называется *эффективной*. Если произведение ковариационной матрицы K_φ состоятельной оценки $\hat{\Theta} = \varphi(U)$ на информационную матрицу Фишера K_z стремится к еди-

ничной матрице при неограниченном увеличении числа опытов n, то оценка $\hat{\Theta}=\varphi(U)$ называется *асимптотически эффективной*.

Из приведенного доказательства соотношения (13) следует, что знак равенства в (13) будет тогда и только тогда, когда $\Gamma_v=0$ и, следовательно, с вероятностью единица $V=0$, т. е.

$$\mu_\varphi^T Z=\mu_\varphi^T \frac{\partial \ln g(U \mid \theta)}{\partial \theta}=\Gamma_\varphi^{-1}(\hat{\Theta}-\theta). \tag{15}$$

Отсюда, принимая во внимание, что согласно (4) $MZ=0$, заключаем, что математическое ожидание эффективной оценки $\hat{\Theta}$ равно θ, т. е. что *любая эффективная оценка несмещенна*. Но в случае несмещенной оценки $\Gamma_\varphi=K_\varphi$, $m'_\varphi=\mu_\varphi=I$ и в силу (14) $K_\varphi^{-1}=K_z$. Следовательно, (15) можно записать в виде

$$Z=\frac{\partial \ln g(U \mid \theta)}{\partial \theta}=K_z(\hat{\Theta}-\theta)=K_z[\varphi(U)-\theta]. \tag{16}$$

Это условие необходимо и достаточно для того, чтобы $\hat{\Theta}=\varphi(U)$ было эффективной оценкой векторного параметра θ.

1.6. Нижние грани дисперсий координат оценки векторного параметра. Из (16) ясно, что *если эффективная оценка векторного параметра существует, то она единственна*.

Из сравнения (16) с условием (9) эффективности скалярной оценки следует, что *координаты эффективной оценки векторного параметра* θ *являются эффективными оценками соответствующих координат вектора* θ *тогда и только тогда, когда ковариационная матрица* K_z *вектора* Z *диагональна*.

► Из доказанной неотрицательной определенности матрицы $K_z-K_\varphi^{-1}$ вытекает неотрицательная определенность матрицы $K_\varphi-K_z^{-1}$ *). Отсюда следует, что диагональные элементы матрицы K_φ не могут быть меньше соответствующих диагональных элементов матрицы K_z^{-1}. Это дает следующую нижнюю границу для дисперсий оценок координат вектора θ:

$$D\hat{\Theta}_p \geqslant K_{pp}^z / |K_z|, \tag{17}$$

*) Для доказательства достаточно привести обе матрицы K_z и K_φ^{-1} к диагональной форме одним и тем же линейным преобразованием.

где K_{pp}^z — алгебраическое дополнение элемента $k_{pp}^z = DZ_p = M|\partial \ln g/\partial\theta_p|^2$ в определителе $|K_z|$ матрицы K_z. Эта нижняя граница достигается для координат эффективной оценки $\hat{\Theta}$ вектора θ.

Заметим, наконец, что так как в случае эффективной оценки $\hat{\Theta}$ оценка $\hat{\Theta}_p$ параметра θ_p является эффективной только в случае диагональной матрицы K_z, то из полученных результатов вытекает неравенство

$$K_{pp}^z / |K_z| \geqslant 1/k_{pp}^z,$$

причем знак равенства имеет место тогда и только тогда, когда матрица K_z диагональна. ◀

Таким образом, в случае оценки неизвестного векторного параметра θ нижняя грань (17) дисперсии координаты $\hat{\Theta}_p$ его несмещенной оценки $\hat{\Theta}$ может быть больше нижней грани (8) дисперсии несмещенной оценки $\hat{\Theta}_p$ параметра θ_p в случае, когда θ_p является единственным неизвестным скалярным параметром. Объясняется это тем, что во втором случае оценками могут служить функции результатов наблюдений U, зависящие от известных остальных координат вектора θ, вследствие чего класс допустимых оценок во втором случае шире, чем в первом.

1.7. Достаточность эффективной оценки. Докажем, что эффективная оценка всегда достаточна.

▶ Пусть $\hat{\Theta} = \varphi(U)$ — любая оценка r-мерного векторного параметра θ, $\varphi_{r+1}(U), \ldots, \varphi_{mn}(U)$ — любые скалярные статистики такие, что при всех реализациях u матрицы U якобиан функций $\varphi_1(u), \ldots, \varphi_r(u), \varphi_{r+1}(u), \ldots, \varphi_{mn}(u)$ по элементам u_{pq} ($p = 1, \ldots, m$; $q = 1, \ldots, n$) отличен от нуля и конечен:

$$J = \frac{\partial(\varphi_1, \ldots, \varphi_{mn})}{\partial(u_{11}, \ldots, u_{mn})} \neq 0, \qquad |J| < \infty.$$

Тогда, вводя векторную статистику Y с координатами

$$Y_p = \varphi_{r+p}(U) \qquad (p = 1, \ldots, mn - r),$$

можно выразить плотность результатов наблюдений U через совместную плотность $h(\hat{\theta}, y|\theta)$ оценки $\hat{\Theta}$ и статистики Y формулой (5.31):

$$g(u|\theta) = h(\hat{\theta}, y|\theta)|J|.$$

Но $h(\hat{\theta}, y|\theta) = p(\hat{\theta}|\theta)\,q(y|\hat{\theta}, \theta)$, где $p(\hat{\theta}|\theta)$ — плотность оценки $\hat{\Theta}$, а $q(y|\hat{\theta}, \theta)$ — условная плотность статистики Y.

Следовательно,

$$g(u\,|\,\theta)=p(\hat{\theta}\,|\,\theta)\,q(y\,|\,\hat{\theta},\ \theta)\,|\cdot J\,|.$$

Так как якобиан J не зависит от θ, то отсюда следует, что

$$\frac{\partial \ln g(u\,|\,\theta)}{\partial\theta}=\frac{\partial \ln p(\hat{\theta}\,|\,\theta)}{\partial\theta}+\frac{\partial \ln q(y\,|\,\hat{\theta},\ \theta)}{\partial\theta}. \tag{18}$$

Отсюда и из (16) для случая эффективной оценки $\hat{\Theta}$ получаем

$$\frac{\partial \ln p(\hat{\theta}\,|\,\theta)}{\partial\theta}=K_z(\hat{\theta}-\theta)-\frac{\partial \ln q(y\,|\,\hat{\theta},\ \theta)}{\partial\theta}.$$

Левая часть не зависит от y. Следовательно, и правая часть не зависит от y. Поэтому

$$\int_{-\infty}^{\infty}\frac{\partial \ln q(y\,|\,\hat{\theta},\ \theta)}{\partial\theta}\,q(y\,|\,\hat{\theta},\ \theta)\,dy=\frac{\partial \ln q(y\,|\,\hat{\theta},\ \theta)}{\partial\theta}.$$

Но для любой плотности, в том числе и для q, справедлива формула (3). Следовательно, $\partial \ln q(y\,|\,\hat{\theta},\ \theta)/\partial\theta=0$ и

$$\frac{\partial \ln p(\hat{\theta}\,|\,\theta)}{\partial\theta}=K_z(\hat{\theta}-\theta). \tag{19}$$

Плотность $q(y\,|\,\hat{\theta},\ \theta)$ в этом случае не зависит от параметра θ ни при каком выборе функций $\varphi_{r+1},\ \ldots,\ \varphi_{mn}$, что и доказывает достаточность эффективной оценки. ◀

Таким образом, *для того чтобы оценка* Θ *была эффективной, необходимо и достаточно, чтобы она была достаточной и ее плотность удовлетворяла условию* (19).

1.8. Случай независимых опытов. В случае независимых опытов плотность $g(u\,|\,\theta)$ равна произведению плотностей значений $X_1,\ \ldots,\ X_n$ случайной величины X, полученных в результате опытов. Поэтому, обозначив плотность величины X через $f(x\,|\,\theta)$, будем иметь

$$\ln g(u\,|\,\theta)=\sum_{k=1}^{n}\ln f(x_k\,|\,\theta).$$

Отсюда, вновь пользуясь независимостью случайных величин $X_1,\ \ldots,\ X_n$ и принимая во внимание (4), получаем

$$DZ=\sum_{k=1}^{n}D\,\frac{\partial \ln f(X_k\,|\,\theta)}{\partial\theta}=nD\,\frac{\partial \ln f(X\,|\,\theta)}{\partial\theta}$$

в случае скалярного θ и

$$K_z = \sum_{p,\,q=1}^{n} M \frac{\partial \ln f(X_p \mid \theta)}{\partial \theta} \frac{\partial^T \ln f(X_q \mid \theta)}{\partial \theta} =$$

$$= \sum_{p=1}^{n} M \frac{\partial \ln f(X_p \mid \theta)}{\partial \theta} \frac{\partial^T \ln f(X_p \mid \theta)}{\partial \theta} = nK,$$

где $K = M[(\partial \ln f/\partial\theta)(\partial^T \ln f/\partial\theta)]$ — ковариационная матрица случайного вектора

$$W = \frac{\partial \ln f(X \mid \theta)}{\partial \theta},$$

в случае векторного θ. Вследствие этого неравенства (7) и (8) в случае скалярного θ принимают вид

$$M(\hat{\Theta} - \theta)^2 \geqslant \frac{[m'_\varphi(\theta)]^2}{nD\,\partial \ln f/\partial\theta}, \tag{20}$$

$$D\hat{\Theta} \geqslant \frac{1}{nD\,\partial \ln f/\partial\theta}, \tag{21}$$

а условия (9) и (16) эффективности оценки $\hat{\Theta} = \varphi(U)$ запишутся в виде

$$\sum_{k=1}^{n} \frac{\partial \ln f(x_k \mid \theta)}{\partial \theta} = c\,[\varphi(x_1, \ldots, x_n) - \theta], \tag{22}$$

где $c = nK$, $K = D\,\partial \ln f/\partial\theta$ в случае скалярного параметра θ. Таким образом, эффективная оценка $\Theta = \varphi(X_1, \ldots, X_n)$ существует тогда и только тогда, когда плотность $f(x \mid \theta)$ наблюдаемой величины X при всех $x_1, \ldots, x_n$ удовлетворяет соотношению (22) при некоторой функции φ.

Пример 1. В случае оценки математического ожидания m и дисперсии D нормально распределенной случайной величины X

$$\ln f(x \mid m, D) = -\frac{1}{2}\ln(2\pi D) - \frac{(x-m)^2}{2D}$$

и, следовательно,

$$\frac{\partial \ln f}{\partial m} = \frac{x-m}{D}, \quad \frac{\partial \ln f}{\partial D} = \frac{1}{2D^2}[(x-m)^2 - D],$$

$$\sum_{k=1}^{n} \frac{\partial \ln f(x_k \mid m, D)}{\partial m} = \frac{1}{D}\left(\sum_{k=1}^{n} x_k - nm\right) = \frac{n}{D}(\bar{x} - m),$$

$$\sum_{k=1}^{n} \frac{\partial \ln f(x_k \mid m, D)}{\partial D} = \frac{n}{2D^2}\left[\frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2 - D\right].$$

Из этих равенств следует, что выборочное среднее $\overline{X}$ представляет собой эффективную оценку математического ожидания нормально распределенной случайной величины, а эффективной оценки дисперсии и эффективной оценки векторного параметра (m, D) не существует, так как условие (22) удовлетворяется только для одной координаты m этого параметра *). Так как

$$D\frac{\partial \ln f}{\partial m}=\frac{1}{D}, \qquad D\frac{\partial \ln f}{\partial D}=\frac{1}{2D^2},$$

то правая часть неравенства (21) равна D/n для оценки m и $2D^2/n$ для оценки D. Первая из этих граней, как видно из (6.13), совпадает с дисперсией $\overline{X}$, как и следовало ожидать для эффективной оценки. Сравнивая вторую из этих граней с дисперсией оценки $\hat{D}$ (6.19), видим, что эффективность оценки $\hat{D}$ равна $(n-1)/n$. Так как $(n-1)/n \longrightarrow 1$ при $n \longrightarrow \infty$, то оценка $\hat{D}$ асимптотически эффективна.

Так как

$$\sum_{k=1}^{n}(x_k-m)^2=\sum_{k=1}^{n}(x_k-\overline{x})^2+n(\overline{x}-m)^2=n(\overline{x}-m)^2+s,$$

то совместная плотность результатов опытов выражается в виде

$$g(x_1, \ldots, x_n \mid m, D)=(2\pi D)^{-n/2}\exp\left\{-\frac{n}{2D}(\overline{x}-m)^2-\frac{s}{2D}\right\},$$

т. е. зависит только от значений статистик $\overline{X}$ и S и не зависит непосредственно от $x_1, \ldots, x_n$. На основании последней теоремы п. 6.2.2 отсюда следует, что векторная статистика $(\overline{X}, S)$ достаточна для параметров m, D. По доказанному в п. 6.2.2 и любая двумерная функция величин $\overline{X}$, S является достаточной статистикой для m и D. В частности, вектор $[\overline{X}\ \hat{D}]^T$, так же как и $[\overline{X}\ D^*]^T$, представляет собой достаточную оценку векторного параметра $[m\ D]^T$.

Пример 2. Совершенно так же доказывается, что в случае оценки математического ожидания m и ковариационной матрицы K нормально распределенного вектора X выборочное среднее $\overline{X}$ является единственной эффективной оценкой вектора m, $\hat{K}=S/(n-1)$ — асимптотически эффективной оценкой матрицы K и что пара $\overline{X}$, S является достаточной статистикой для m, K. Из последнего факта вытекает, что пара $\overline{X}$, $\hat{K}$, так же как и $\overline{X}$, K^*, представляет собой достаточную оценку для m, K.

1.9. Случай дискретной наблюдаемой величины. Рассмотрим теперь частный случай, когда наблюдаемая случайная величина X дискретна. Обозначим через $x^{(1)}, \ldots, x^{(N)}$ ее возможные значения, а через $p_1(\theta), \ldots, p_N(\theta)$ их вероятности. В этом случае плотность

*) Если математическое ожидание m известно, то оценка $\tilde{D}=\sum(X_k-m)^2/n$ дисперсии D эффективна.

$f(x\mid\theta)$ определяется формулой

$$f(x\mid\theta)=\sum_{\nu=1}^{N} p_\nu(\theta)\,\delta(x-x^{(\nu)}). \qquad (23)$$

Если значения $x^{(1)}, \ldots, x^{(N)}$ величины X не зависят от неизвестного параметра θ, а вероятности $p_1(\theta), \ldots, p_N(\theta)$ дифференцируемы по θ, то

$$\frac{\partial f(x\mid\theta)}{\partial\theta}=\sum_{\nu=1}^{N}\frac{\partial p_\nu(\theta)}{\partial\theta}\,\delta(x-x^{(\nu)}). \qquad (24)$$

Отсюда видно, что отношение $(\partial f/\partial\theta)/f$ имеет значения $(\partial p_\nu/\partial\theta)/p_\nu$ в точках $x=x^{(\nu)}$ $(\nu=1, \ldots, N)$ и остается неопределенным во всех других точках. Поэтому

$$K=M\frac{\partial \ln f}{\partial\theta}\frac{\partial^T \ln f}{\partial\theta}=\sum_{\nu=1}^{N} p_\nu\frac{\partial \ln p_\nu}{\partial\theta}\frac{\partial^T \ln p_\nu}{\partial\theta}\ ^{*}). \qquad (25)$$

Обозначим через M_ν число появлений значения $x^{(\nu)}$ $(\nu=1, \ldots, N)$ величины X при n опытах. Из (22) следует, что эффективная оценка $\hat{\Theta}=\varphi(M_1, \ldots, M_N)$ параметра θ существует тогда и только тогда, когда при любых целых неотрицательных числах $m_1, \ldots, m_N$, $m_1+\ldots+m_N=n$,

$$\sum_{\nu=1}^{N} m_\nu\frac{\partial \ln p_\nu}{\partial\theta}=c\,[\varphi(m_1, \ldots, m_N)-\theta]. \qquad (26)$$

---

*) Для строгого вывода этой формулы выберем настолько малое $\varepsilon>0$, чтобы каждый интервал $(x^{(\nu)}-\varepsilon,\ x^{(\nu)}+\varepsilon)$ не содержал возможных значений величины X, отличных от $x^{(\nu)}$, и рассмотрим случай, когда плотность $f(x\mid\theta)$ величины X определяется формулой

$$f(x\mid\theta)=\sum_{\nu=1}^{N} p_\nu(\theta)\,\delta_\varepsilon(x-x^{(\nu)}),$$

где $\delta_\varepsilon(x)=1/2\varepsilon$ при $|x|<\varepsilon$ и $\delta_\varepsilon(x)=0$ при $|x|>\varepsilon$. Тогда получим

$$\frac{\partial \ln f}{\partial\theta}=\frac{\partial \ln p_\nu}{\partial\theta} \quad \text{при} \quad |x-x^{(\nu)}|<\varepsilon \quad (\nu=1, \ldots, N)$$

и

$$K=\int_{-\infty}^{\infty}\frac{\partial \ln f(x\mid\theta)}{\partial\theta}\frac{\partial^T \ln f(x\mid\theta)}{\partial\theta}f(x\mid\theta)\,dx=\sum_{\nu=1}^{N} p_\nu\frac{\partial \ln p_\nu}{\partial\theta_\nu}\frac{\partial^T \ln p_\nu}{\partial\theta_\nu}.$$

А так как это справедливо при любом $\varepsilon>0$, то (25) справедливо и в пределе при $\varepsilon\to 0$.

Пример 3. В случае оценки вероятности события в каждом опыте наблюдается число появлений этого события — дискретная случайная величина с двумя возможными значениями 0 и 1, вероятности которых равны соответственно $q=1-p$ и p. Неизвестным параметром служит p. В данном случае

$$\frac{\partial \ln p}{\partial p}=\frac{1}{p}, \quad \frac{\partial \ln q}{\partial p}=-\frac{1}{1-p}=-\frac{1}{q}$$

и при любом числе появлений m события

$$m\frac{\partial \ln p}{\partial p}+(n-m)\frac{\partial \ln q}{\partial p}=\frac{m}{p}-\frac{n-m}{q}=\frac{n}{pq}\left(\frac{m}{n}-p\right).$$

Таким образом, условие (26) удовлетворяется, и, следовательно, существует эффективная оценка вероятности p, представляющая собой частоту события. Коэффициент c в данном случае равен n/pq. Формула (25) дает

$$K=D\frac{\partial \ln f}{\partial p}=\frac{p}{p^2}+\frac{q}{q^2}=\frac{1}{pq},$$

и неравенство (20) принимает вид

$$M(\hat{P}-p)^2 \geqslant \frac{[m'_\varphi(p)]^2}{n/pq}=[m'_\varphi(p)]^2\frac{pq}{n},$$

где $m_\varphi(p)=M\hat{P}$. Из (6.7) следует, что здесь получается знак равенства, если за оценку $\hat{P}$ вероятности p принять частоту. Это еще раз подтверждает, что частота является эффективной оценкой вероятности события.

Пример 4. В случае оценивания вероятностей $p_1, \ldots, p_N$ значений $x_1, \ldots, x_N$ дискретной случайной величины X $\theta=[p_1\ldots p_{N-1}]^T$, $p_N=1-p_1-\ldots-p_{N-1}$, и, следовательно,

$$\frac{\partial \ln p_\nu}{\partial p_\mu}=\frac{\delta_{\nu\mu}}{p_\nu}, \quad \frac{\partial \ln p_N}{\partial p_\mu}=-\frac{1}{p_N} \quad (\mu=1, \ldots, N-1),$$

$$\sum_{\nu=1}^{N} m_\nu\frac{\partial \ln p_\nu}{\partial p_\mu}=\frac{m_\mu}{p_\mu}-\frac{m_N}{p_N}=\frac{n}{p_\mu}\left(\frac{m_\mu}{n}-p_\mu\right)+\frac{n}{p_N}\sum_{\nu=1}^{N-1}\left(\frac{m_\nu}{n}-p_\nu\right).$$

Сравнив это равенство с (26), приходим к заключению, что вектор частот $[\hat{P}_1\ldots\hat{P}_{N-1}]^T$ представляет собой эффективную оценку векторного параметра $[p_1\ldots p_{N-1}]^T$. Так как матрица

$$K=\sum_{\nu=1}^{N} p_\nu\frac{\partial \ln p_\nu}{\partial\theta}\frac{\partial^T \ln p_\nu}{\partial\theta}=$$

$$=\begin{bmatrix} p_1^{-1}+p_N^{-1} & p_N^{-1}\ldots & p_N^{-1} \\ p_N^{-1} & p_2^{-1}+p_N^{-1}\ldots & p_N^{-1} \\ \cdots & \cdots & \cdots \\ p_N^{-1} & p_N^{-1}\ldots p_{N-1}^{-1} & +p_N^{-1} \end{bmatrix}$$

в данном случае не диагональна, то ни одна из частот $\hat{P}_1\ldots\hat{P}_N$ по отдельности не является эффективной оценкой соответствующей вероят-

ности. Это объясняется тем, что в случае одной неизвестной вероятности p_μ при известных остальных можно построить несмещенную оценку, зависящую от известных вероятностей, обладающую меньшей дисперсией, чем частота.

§ 2. Основные методы нахождения оценок

2.1. Метод максимума правдоподобия. Наиболее важным с теоретической точки зрения общим методом нахождения оценок является *метод максимума правдоподобия* Фишера. Этот метод основан на принятии за оценку неизвестного параметра θ такого его значения, при котором плотность результатов опытов $X_1, \ldots, X_n$ достигает наибольшего значения при полученных реализациях $x_1, \ldots$ $\ldots, x_n$ результатов опытов $X_1, \ldots, X_n$:

$$g(u \mid \theta) = g(x_1, \ldots, x_n \mid \theta) = \max. \tag{27}$$

Функция $g(x_1, \ldots, x_n \mid \theta)$ называется *функцией правдоподобия*, а значение θ, при котором она достигает наибольшего значения,— *оценкой максимального правдоподобия* параметра θ. Оценки максимального правдоподобия обычно определяют, максимизируя $\ln g(x_1, \ldots, x_n \mid \theta)$, пользуясь тем, что $y = \ln x$ — строго возрастающая функция.

В случае независимых опытов

$$\ln g(x_1, \ldots, x_n \mid \theta) = \sum_{k=1}^{n} \ln f(x_k \mid \theta), \tag{28}$$

где $f(x \mid \theta)$ — плотность наблюдаемой случайной величины X. Если $f(x \mid \theta)$ при всех x имеет непрерывную производную по θ (непрерывный градиент в случае векторного θ), то для определения оценки максимального правдоподобия получаем уравнение

$$\sum_{k=0}^{n} \frac{\partial \ln f(x_k \mid \theta)}{\partial \theta} = 0. \tag{29}$$

В случае r-мерного векторного параметра θ это уравнение представляет собой векторное уравнение той же размерности, т. е. систему r скалярных уравнений относительно координат вектора θ. Естественно, что оценками параметра θ могут служить только те решения уравнения максимума правдоподобия (29), которые зависят от результатов опытов $x_1, \ldots, x_n$. Если уравнение (29) имеет решения, не зависящие от $x_1, \ldots, x_n$, то их следует отбросить.

Если существует эффективная оценка $\hat{\theta} = \varphi(x_1, \ldots, x_n)$, то на основании (22) уравнение максимума правдоподобия (29) принимает вид

$$c[\varphi(x_1, \ldots, x_n) - \theta] = 0.$$

Отсюда следует, что *если существует эффективная оценка параметра* θ, *то она является единственным решением уравнения максимума правдоподобия.*

Предположим теперь, что оценка $\hat{\theta} = \varphi(x_1, \ldots, x_n)$ достаточна для θ. В этом случае плотность $q(y | \hat{\theta}, \theta)$ в (18) не зависит от θ и на основании (27), (28) и (18) уравнение максимума правдоподобия (29) принимает вид

$$\frac{\partial \ln p(\hat{\theta} | \theta)}{\partial \theta} = 0. \tag{30}$$

Отсюда видно, что *в случае существования достаточной оценки (или статистики) для* θ *все решения уравнения максимума правдоподобия являются функциями этой достаточной оценки (статистики).* По доказанному в п. 6.2.2 отсюда следует, что *если существует достаточная статистика для* θ, *то все решения уравнения максимума правдоподобия являются достаточными оценками параметра* θ.

В случае дискретной наблюдаемой случайной величины X с возможными значениями $x^{(1)}, \ldots, x^{(N)}$, вероятности которых равны $p_1(\theta), \ldots, p_N(\theta)$, функция $\partial \ln f(x | \theta)/\partial \theta$ имеет значение $\partial \ln p_\nu(\theta)/\partial \theta$ в точке $x^{(\nu)}$ $(\nu = 1, \ldots, N)$. Поэтому, если m_ν из экспериментальных значений $x_1, \ldots$ $\ldots, x_n$ наблюдаемой величины X равны $x^{(\nu)}$ $(\nu = 1, \ldots, N)$, $m_1 + \ldots + m_N = n$, то уравнение максимума правдоподобия (29) имеет вид

$$\sum_{\nu=1}^{N} m_\nu \frac{\partial \ln p_\nu(\theta)}{\partial \theta} = 0. \tag{31}$$

Оценки максимального правдоподобия при достаточно общих условиях состоятельны, асимптотически нормальны и асимптотически эффективны [35, 45, 86].

Пример 5. В случае оценки вероятности события p уравнение (31) принимает вид (пример 3)

$$\frac{m}{p} - \frac{n-m}{1-p} = \frac{n}{pq}\left(\frac{m}{n} - p\right) = 0,$$

где m — число появлений события. Единственным решением этого уравнения является частота, $p = m/n$, в полном согласии с тем, что частота представляет собой эффективную оценку вероятности p.

Пример 6. В случае оценки математического ожидания m и дисперсии D нормально распределенной случайной величины X на основании результатов примера 1 уравнение максимума правдоподобия (29) представляет собой систему двух уравнений

$$\frac{n}{D}(\bar{x} - m) = 0, \quad \frac{n}{2D^2}\left[\frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2 - D\right] = 0.$$

Первое из этих уравнений имеет единственное решение $m = \bar{x}$ ($\bar{X}$ — эффективная оценка). Подставив это значение m во второе уравнение и решив его, получаем

$$D = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2.$$

Таким образом, метод максимума правдоподобия дает в данном случае выборочное среднее $\bar{X}$ в качестве оценки математического ожидания m и смещенную оценку D^* для дисперсии D.

Пример 7. В случае оценки математического ожидания m и ковариационной матрицы K нормально распределенного r-мерного случайного вектора X

$$\ln f(x \mid m, K) = \frac{1}{2}\ln|C| - \frac{n}{2}\ln(2\pi) - \frac{1}{2}(x^T - m^T)C(x - m) =$$

$$= \frac{1}{2}\ln|C| - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{p,q=1}^{r} c_{pq}(x_p - m_p)(x_q - m_q),$$

где $C = K^{-1}$. Так как матрицы C и K связаны взаимно однозначной зависимостью, то при составлении уравнений максимума правдоподобия можно взять производные $\ln f(x \mid m, K)$ по элементам матрицы C. В результате, принимая во внимание, что производная определителя по любому его элементу равна алгебраическому дополнению этого элемента, находим

$$\frac{\partial \ln f(x \mid m, K)}{\partial m_p} = \sum_{s=1}^{r} c_{ps}(x_s - m_s),$$

$$\frac{\partial \ln f(x \mid m, K)}{\partial c_{pq}} = \frac{C_{pq}}{2|C|} - \frac{1}{2}(x_p - m_p)(x_q - m_q) =$$

$$= \frac{1}{2}[k_{pq} - (x_p - m_p)(x_q - m_q)],$$

где C_{pq} — алгебраическое дополнение элемента c_{pq} в $|C|$. Подставляя сюда вместо x, x_p соответственно x_k, x_{kp}, суммируя полученные выражения по k и приравнивая результаты нулю, получим систему

уравнений максимума правдоподобия

$$\sum_{s=1}^{r} c_{sp}\left(\sum_{k=1}^{n} x_{ks}-nm_s\right)=0,$$

$$nk_{pq}-\sum_{k=1}^{n}(x_{kp}-m_p)(x_{kq}-m_q)=0 \qquad (p,\ q=1,\ \ldots,\ r).$$

Так как определитель матрицы C отличен от нуля, то первая группа уравнений имеет единственное решение

$$m_s=\frac{1}{n}\sum_{k=1}^{n} x_{ks}=\bar{x}_s \qquad (s=1,\ \ldots,\ r).$$

Подставив это решение во вторую группу уравнений, находим

$$k_{pq}=\frac{1}{n}\sum_{k=1}^{n}(x_{kp}-\bar{x}_p)(x_{kq}-\bar{x}_q) \qquad (p,\ q=1,\ \ldots,\ r).$$

Таким образом, метод максимума правдоподобия дает в данном случае выборочное среднее $\bar{X}$ как оценку математического ожидания и смещенную оценку $K^*=S/n=(n-1)\hat{K}/n$ для ковариационной матрицы.

Пример 8. Показать, что если наблюдаемая случайная величина X имеет показательное распределение $f(x\mid\lambda)=\lambda e^{-\lambda x}1(x)$, то оценкой максимума правдоподобия параметра λ служит величина, обратная выборочному среднему значению величины $\bar{X}$: $\hat{\Lambda}=1/\bar{X}$. Эффективна эта оценка или нет? Достаточна ли она? Решить те же вопросы в случае, когда оценивается параметр $\nu=\lambda^{-1}=MX$.

Пример 9. В случае оценки параметров λ и μ в γ-распределении

$$f(x\mid\lambda,\ \mu)=\frac{\lambda^{\mu+1}x^{\mu}}{\Gamma(\mu+1)}e^{-\lambda x}1(x)$$

уравнения максимума правдоподобия имеют вид

$$n\left(\frac{\mu+1}{\lambda}-\bar{x}\right)=0,$$

$$n\ln\lambda+\sum_{k=1}^{n}\ln x_k-n\frac{\Gamma'(\mu+1)}{\Gamma(\mu+1)}=0.$$

Первое из этих уравнений дает

$$\lambda=(\mu+1)/\bar{x}.$$

Подставив это выражение во второе уравнение, приводим его к виду

$$(\mu+1)e^{-\Psi(\mu)}=\frac{\bar{x}}{\sqrt[n]{x_1\ldots x_n}}, \qquad \Psi(\mu)=\frac{\Gamma'(\mu+1)}{\Gamma(\mu+1)}.$$

Это уравнение можно решить методом последовательных приближений, пользуясь таблицей функции $\Psi(z)$. После этого найдем оценку

параметра λ. Показать, что статистики $\overline{X}$ и $S=\sqrt[n]{X_1\ldots X_n}$ достаточны и что эффективных оценок параметров λ и μ не существует. Обратите внимание на то, что $\overline{X}$ служит эффективной оценкой параметра $\nu=(\mu+1)\lambda^{-1}=MX$, так же как и в случае показательного распределения ($\mu=0$).

Пример 10. В случае оценки параметра $\mu=MX$ в распределении Пуассона $p_\nu(\mu)=\mu^\nu e^{-\mu}/\nu!$ ($\nu=0, 1, 2\ldots$) и уравнение максимума правдоподобия (31) принимает вид

$$\frac{n}{\mu}\left(\frac{1}{n}\sum_{\nu=0}^{\infty}\nu m_\nu-\mu\right)=0.$$

Это уравнение имеет единственное решение

$$\mu=\frac{1}{n}\sum_{\nu=0}^{\infty}\nu m_\nu,$$

представляющее собой эффективную оценку параметра μ. Но

$$\sum_{\nu=0}^{\infty}\nu m_\nu=\sum_{k=1}^{n}x_k,$$

так как m_ν представляет собой число экспериментальных значений величины X, равных ν, т. е. число тех x_k, которые равны ν. Таким образом, и в этом случае выборочное среднее служит эффективной оценкой математического ожидания случайной величины.

Пример 11. Во всех рассмотренных примерах выборочное среднее оказалось эффективной оценкой математического ожидания. Чтобы показать, что это не всегда так, достаточно рассмотреть случай, когда эффективной оценки математического ожидания заведомо не существует, например упомянутый в § 1 случай разрывной плотности с зависящими от математического ожидания точками разрыва, когда правая часть неравенства (7) равна 0. В качестве такого примера рассмотрим случайную величину X, равномерно распределенную на интервале $(m-\alpha,\ m+\alpha)$ с неизвестным параметром $m=MX$. В этом случае $f(x\mid m)=1/2\alpha$ при $x\in(m-\alpha,\ m+\alpha)$ и функция правдоподобия постоянна и равна $1/(2\alpha)^n$ при всех значениях m, для которых $x_k\in(m-\alpha,\ m+\alpha)$ $(k=1,\ldots,n)$, и равна нулю при всех остальных m. Но $x_k\in(m-\alpha,\ m+\alpha)$ при $m\in(x_k-\alpha,\ x_k+\alpha)$. Следовательно, множеством значений m, при которых функция правдоподобия отлична от нуля, служит пересечение всех интервалов $(x_k-\alpha,\ x_k+\alpha)$ $(k=1,\ldots$ $\ldots, n)$, т. е. интервал $(\max x_k-\alpha,\ \min x_k+\alpha)$. Любое значение m в этом интервале представляет собой оценку максимума правдоподобия математического ожидания m. При этом выборочное среднее $\overline{x}$ может и не принадлежать этому интервалу.

Пример 12. Если надо оценить оба параметра равномерного распределения m и α, то, поскольку функция правдоподобия обратно пропорциональна α^n, следует взять в качестве оценки α минимальное значение α, совместимое с результатами опытов, т. е. минимальное α, при котором $\max x_k-\alpha\leqslant\min x_k+\alpha$. Таким значением α будет $(\max x_k-\min x_k)/2$. При этом для m получаем единственную оценку $m=(\max x_k+\min x_k)/2$.

2.2. Одно свойство метода максимума правдоподобия. Для нахождения оценок максимума правдоподобия полезна следующая теорема: *если параметры* θ *и* ϑ *связаны взаимно однозначной непрерывной зависимостью, то их оценки максимума правдоподобия связаны той же зависимостью.*

► Предположим, что параметры θ и ϑ связаны взаимно однозначной зависимостью $\vartheta = \varphi(\theta)$, $\theta = \varphi^{-1}(\vartheta)$, причем функции φ и φ^{-1} непрерывны. Пусть $\hat{\theta}$ — оценка максимума правдоподобия параметра θ и $\hat{\vartheta} = \varphi(\hat{\theta})$. Требуется доказать, что $\hat{\vartheta}$ — оценка максимума правдоподобия параметра ϑ.

Из того, что $\hat{\theta}$ — оценка максимума правдоподобия, следует, что $g(u \mid \theta) < g(u \mid \hat{\theta})$ в некоторой окрестности точки $\hat{\theta}$. Отсюда в силу соотношений $\theta = \varphi^{-1}(\vartheta)$, $\hat{\theta} = \varphi^{-1}(\hat{\vartheta})$ и непрерывности φ^{-1} следует, что $g(u \mid \varphi^{-1}(\vartheta)) < g(u \mid \varphi^{-1}(\hat{\vartheta}))$ в некоторой окрестности точки $\hat{\vartheta}$.

Если $\hat{\theta}$ — единственная оценка максимума правдоподобия параметра θ, то $g(u \mid \theta) < g(u \mid \hat{\theta})$ при всех θ и, следовательно, $g(u \mid \varphi^{-1}(\vartheta)) < g(u \mid \varphi^{-1}(\hat{\vartheta}))$ при всех ϑ, т. е. $\hat{\vartheta}$ — единственная оценка максимума правдоподобия параметра ϑ. ◄

Пример 13. Найти оценки максимума правдоподобия собственных значений и собственных векторов ковариационной матрицы K_x r-мерного случайного вектора X.

В данном случае роль параметра θ играет K_x, а роль параметра ϑ — вектор, координатами которого служат собственные значения $\lambda_1, \ldots, \lambda_r$ и независимые координаты собственных векторов $\varphi_1, \ldots, \varphi_r$. Покажем, что векторы θ и ϑ имеют одну и ту же размерность. В силу симметрии ковариационной матрицы она полностью определяется $r(r+1)/2$ своими элементами k_{pq} при $q \geqslant p$. С другой стороны, r^2 координат векторов $\varphi_1, \ldots, \varphi_r$ связаны $r(r+1)/2$ соотношениями (3.62), $\varphi_p^T \varphi_q = \delta_{pq}$, $q \geqslant p$. Поэтому число независимых координат векторов $\varphi_1, \ldots, \varphi_r$ равно $r^2 - r(r+1)/2$. Таким образом, собственные значения $\lambda_1, \ldots, \lambda_r$ и собственные векторы $\varphi_1, \ldots, \varphi_r$ полностью определяются $r + r^2 - r(r+1)/2 = r(r+1)/2$ числами. Далее, в силу соотношений (3.61), (3.67) и (3.68) в случае различных собственных значений $\lambda_1, \ldots, \lambda_r$ матрица K_x и совокупность собственных значений $\lambda_1, \ldots, \lambda_r$ и собственных векторов $\varphi_1, \ldots, \varphi_r$ связаны взаимно однозначной непрерывной зависимостью. По доказанной теореме оценки максимума правдоподобия параметров K_x, с одной стороны, и $\lambda_1, \ldots, \lambda_r$, $\varphi_1, \ldots, \varphi_r$, с другой стороны, связаны той же зависимостью. Следовательно, оценками максимума правдоподобия собственных значений $\lambda_1, \ldots, \lambda_r$ и собственных векторов $\varphi_1, \ldots, \varphi_r$ матрицы K_x служат собственные значения $\lambda_1^*, \ldots, \lambda_r^*$ и собственные векторы $\varphi_1^*, \ldots, \varphi_r^*$

матрицы

$$K_x^* = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})(X_k^T - \overline{X}^T), \qquad \overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$$

(которая согласно результату примера 6 служит оценкой максимума правдоподобия матрицы K_x).

В случае, когда среди собственных значений $\lambda_1, \ldots, \lambda_r$ имеются совпадающие, соответствующие собственные векторы определяются неоднозначно. Однако и в этом случае можно утверждать, что собственные значения и собственные векторы оценки максимального правдоподобия матрицы K_x представляют собой оценки максимального правдоподобия собственных значений и собственных векторов, соответствующих простым собственным значениям, и некоторых соответственно выбранных собственных векторов, соответствующих кратным собственным значениям.

Можно доказать, что в случае различных собственных значений $\lambda_1, \ldots, \lambda_r$ матрицы K_x собственные значения $\lambda_1^*, \ldots, \lambda_r^*$ с вероятностью 1 различны и, следовательно, собственные векторы $\varphi_1^*, \ldots$ $\ldots, \varphi_r^*$ определяются однозначно с вероятностью 1.

2.3. Метод моментов. Вторым общим методом нахождения оценок параметров распределений является *метод моментов*. Этот метод основан на определении неизвестных параметров из уравнений, получаемых приравниванием экспериментально найденных оценок моментов теоретическим значениям соответствующих моментов, зависящим от неизвестных параметров. Так, например, если плотность $f(x \mid \theta)$ наблюдаемой случайной величины X зависит от неизвестного r-мерного векторного параметра θ, то для нахождения оценки этого параметра приравнивают оценки каких-нибудь r моментов величины X соответствующим теоретическим моментам. В случае скалярной наблюдаемой величины X r-мерный параметр θ обычно определяют из уравнений

$$\alpha_p(\theta) = \frac{1}{n} \sum_{k=1}^{n} x_k^p \qquad (p = 1, \ldots, r).$$

В случае векторной наблюдаемой величины X берут аналогичные уравнения для r моментов возможно меньших порядков. Как правило, ограничиваются моментами не выше четвертого порядка.

Метод моментов часто оказывается проще, чем метод максимума правдоподобия.

Пример 14. Оценим параметр λ показательного распределения методом моментов. При этом достаточно оценить один первый момент —

математическое ожидание $MX = 1/\lambda$. В результате получим ту же оценку, что и методом максимума правдоподобия.

Пример 15. В случае оценки параметров λ, μ γ-распределения, приравнивая первый и второй моменты соответствующим оценкам, получаем уравнения

$$\frac{\mu+1}{\lambda} = \frac{1}{n}\sum_{k=1}^{n} x_k, \qquad \frac{(\mu+1)(\mu+2)}{\lambda^2} = \frac{1}{n}\sum_{k=1}^{n} x_k^2$$

(теоретические выражения моментов для γ-распределения были получены в примере 3.16). Разделив второе из этих уравнений на первое, возведенное в квадрат, получим

$$\frac{\mu+2}{\mu+1} = \frac{1}{n\bar{x}^2}\sum_{k=1}^{n} x_k^2.$$

Решив это уравнение, находим

$$\mu = \frac{2x^2 - \frac{1}{n}\sum\limits_{k=1}^{n} x_k^2}{\frac{1}{n}\sum\limits_{k=1}^{n} x_k^2 - \bar{x}^2}.$$

После этого находим оценку параметра λ:

$$\lambda = \frac{\mu+1}{\bar{x}} = \frac{\bar{x}}{\frac{1}{n}\sum\limits_{k=1}^{n} x_k^2 - \bar{x}^2}.$$

Мы видим, что в данном случае оценки параметров методом моментов находятся значительно проще, чем методом максимума правдоподобия в примере 9.

Пример 16. Оценим методом моментов параметры m и α равномерного распределения на интервале $(m-\alpha,\ m+\alpha)$. Пользуясь оценками первых двух моментов и результатами вычисления моментов для равномерного распределения в примере 3.3, получаем уравнения

$$m = \frac{1}{n}\sum_{k=1}^{n} x_k, \qquad \frac{3m^2+\alpha^2}{3} = \frac{1}{n}\sum_{k=1}^{n} x_k^2.$$

Отсюда находим $m = \bar{x}$ и

$$\alpha^2 = \frac{3}{n}\left(\sum_{k=1}^{n} x_k^2 - n\bar{x}^2\right).$$

Таким образом, метод моментов дает в качестве оценки параметра m равномерного распределения выборочное среднее.

Еще один метод нахождения оценок параметров распределений — *метод минимума* χ^2 мы изложим в § 8.3 в связи с задачами оценивания законов распределения.

Наиболее общие методы нахождения и исследования оценок случайных величин и статистических характеристик дает сравнительно новая и бурно развивающаяся отрасль теории вероятностей — статистическая теория решений [15].

§ 3. Рекуррентное оценивание корня уравнения регрессии

3.1. Рекуррентное оценивание математического ожидания. Изложенные методы нахождения оценок неизвестных параметров, от которых зависят распределения наблюдаемых случайных величин, дают возможность оценивать неизвестные параметры только после того, как произведены все опыты. Между тем во многих задачах практики приходится совмещать опыты с решением определенных практических задач, эффективность решения которых существенно зависит от точности приближения к неизвестным параметрам. В таких случаях желательно производить оценку неизвестных параметров после каждого опыта с тем, чтобы повысить эффективность решения задач в следующем опыте. Таковы, например, задачи пристрелки при стрельбе и юстировки (устранения систематических ошибок) точных приборов. Задачей пристрелки является определение отклонения центра рассеивания от цели, т. е. математического ожидания отклонения точки попадания от цели, по результатам измерений отклонений в процессе стрельбы с тем, чтобы по возможности точнее совместить центр рассеивания с целью. Ясно, что для повышения эффективности каждого следующего выстрела желательно вносить поправку после каждого выстрела, а не ждать, когда будут произведены все выстрелы, предназначенные для пристрелки. Таким образом, возникает необходимость *рекуррентного оценивания* неизвестных параметров, когда после каждого опыта определяется поправка к оценке, найденной по результатам предыдущих опытов.

▶ Разъясним основную идею рекуррентного оценивания на простейшем примере оценивания математического ожидания случайной величины. При этом мы будем отмечать выборочное среднее значение случайной величины X, полученное по результатам n опытов, верхним

индексом n. Тогда будем иметь

$$\overline{X}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{n-1}{n}\,\frac{1}{n-1}\sum_{k=1}^{n-1} X_k + \frac{1}{n} X_n =$$
$$= \overline{X}^{(n-1)} + \frac{1}{n}\left(X_n - \overline{X}^{(n-1)}\right).$$

Таким образом, мы получили простую рекуррентную формулу для выборочных средних:

$$\overline{X}^{(n)} = \overline{X}^{(n-1)} + \frac{1}{n}\left(X_n - \overline{X}^{(n-1)}\right). \blacktriangleleft \qquad (32)$$

Эта формула дает простое правило нахождения поправок к ранее найденной оценке математического ожидания: *после первого опыта за оценку математического ожидания принимается наблюденное значение* X_1 *случайной величины,* $\overline{X}^{(1)} = X_1$, *после второго опыта добавляется поправка, равная половине отклонения* $X_2 - \overline{X}^{(1)}$, *и после n-го опыта к найденной ранее оценке* $\overline{X}^{(n-1)}$ *добавляется поправка, равная одной n-й отклонения* $X_n - \overline{X}^{(n-1)}$ (после первого выстрела или измерения вносится поправка, равная наблюдаемому отклонению, после второго — поправка, равная половине наблюденного отклонения, и вообще после n-го — поправка, равная наблюденному отклонению, деленному на n). Это правило издавна применяется в артиллерии и в точной механике.

Очевидно, что рекуррентная формула (32) применима и для оценок любых моментов скалярных и векторных случайных величин, поскольку оценкой любого (начального или относительно данной точки) момента служит среднее арифметическое соответствующей степени случайной величины (произведения степеней координат случайного вектора).

Процесс рекуррентного оценивания, когда поправка после каждого опыта зависит только от результата этого опыта и предыдущей оценки, называется *стохастической аппроксимацией*. Это название объясняется тем, что приближение к неизвестным параметрам достигается последовательностью случайных величин (в нашем примере $\{\overline{X}^{(n)}\}$). Общая теория стохастических аппроксимаций была заложена работой Роббинса и Монро [92] и развита в последующих работах Кифера и Вольфовица [36], Блюма [5, 6] и многих других, хотя, как мы видели,

метод стохастических аппроксимаций был известен и применялся задолго до создания общей теории.

Чтобы сформулировать общую задачу оценивания параметров методом стохастических аппроксимаций, рассмотрим процесс применения формулы (32) с другой точки зрения. Введем случайную величину $Y = X - z$ при произвольном z. Ее математическое ожидание при данном z равно $M[Y \mid z] = MX - z = m_x - z$, т. е. представляет собой функцию переменной z — регрессию случайной величины Y на z (п. 4.3.2). При этом функция $\psi(z) = M[Y \mid z]$ известна не полностью, поскольку она зависит от неизвестного параметра m_x. Задача оценки m_x тогда сведется к оценке корня уравнения регрессии $\psi(z) = 0$. Ввиду состоятельности оценки $\bar{X}^{(n)}$ математического ожидания m_x последовательность случайных величин $Z_n = \bar{X}^{(n)}$ $(n = 1, 2, \ldots)$ сходится п. в. к корню m_x уравнения $\psi(z) = 0$. При этом в n-м опыте наблюдается случайная величина $Y_n = X_n - Z_{n-1}$ $(n = 1, 2, \ldots)$, $Z_0 = 0$, после чего определяется реализация случайной величины

$$Z_n = Z_{n-1} + \frac{1}{n} Y_n. \tag{33}$$

Эта трактовка рассмотренной задачи подводит к широким обобщениям. В данном случае функция регрессии $\psi(z) = m_x - z$ линейна и зависит только от одного неизвестного параметра m_x (векторного в случае векторной величины X). В общем случае $\psi(z) = M[Y \mid z]$ может быть неизвестной функцией, удовлетворяющей лишь некоторым самым общим условиям. Оказывается, что и в таких случаях методом стохастических аппроксимаций можно по результатам последовательных наблюдений величины Y оценивать корень уравнения регрессии $\psi(z) = \alpha$ при любом α, а для скалярной величины Y — и точку экстремума $\psi(z)$.

3.2. Процесс стохастических аппроксимаций. Будем считать параметр z и случайную величину Y, а следовательно, и функцию $\psi(z) = M[Y \mid z]$ в общем случае r-мерными векторами. Предположим, что уравнение $\psi(z) = \alpha$ имеет единственное решение θ и что при $z_q = \theta_q$, $q \neq p$, функция $\omega_p(z_p) = (z_p - \theta_p)[\psi_p(z) - \alpha_p]$ переменной z_p сохраняет постоянный знак в достаточно большой окрестности V_p точки θ_p $(p = 1, \ldots, r)$. Положим $\varepsilon_p = -\operatorname{sgn} \omega_p(z_p)$ и обозначим через ε диагональную матрицу с элементами $\varepsilon_1, \ldots, \varepsilon_r$. Для оценки корня θ уравнения $\psi(z) = \alpha$

возьмем произвольную случайную или неслучайную величину Z_0 (конечно, такую, чтобы вероятность ее попадания в окрестность $V_1 \times \ldots \times V_r$ точки θ была достаточно близка к 1) и построим процесс стохастических аппроксимаций по формуле

$$Z_n = Z_{n-1} + a_n \varepsilon_n (Y_n - \alpha), \tag{34}$$

где Y_n — результат наблюдения случайной величины Y при $z = Z_{n-1}$ $(n = 1, 2, \ldots)$, $\{a_n\}$ — последовательность положительных чисел, удовлетворяющих условиям

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty, \tag{35}$$

а ε_n — значение матрицы ε при $z = Z_{n-1}$. В частности, этим условиям удовлетворяет последовательность $\{1/n\}$, с помощью которой строится последовательность (33), сходящаяся п. в. к неизвестному математическому ожиданию. Смысл условий (35) состоит в следующем: первое условие — расходимость ряда с общим членом $a_n > 0$ — обеспечивает возможность неограниченного приближения Z_n к θ, не дает процессу приближений остановиться, не дойдя до точки θ; второе условие — сходимость ряда с общим членом a_n^2 — обеспечивает ограниченность последовательности дисперсий DZ_n, не дает процессу неограниченно удаляться от точки θ. Диагональная матрица ε с элементами, равными 1, введена в (34) для того, чтобы обеспечить в среднем приближение к θ. Для иллюстрации рассмотрим случай скалярного параметра θ. Из (34) находим

$$M[Z_n | Z_{n-1}] - Z_{n-1} = a_n \varepsilon [\psi(Z_{n-1}) - \alpha] = \\ = -a_n [\psi(Z_{n-1}) - \alpha] \operatorname{sgn} \omega(Z_{n-1}),$$

где $\omega(z) = (z - \theta)[\psi(z) - \alpha]$. Отсюда видно, что знак разности $M[Z_n | Z_{n-1}] - Z_{n-1}$ всегда противоположен знаку отклонения $Z_{n-1} - \theta$, что и обеспечивает в среднем шаг от Z_{n-1} в сторону θ.

3.3. Сходимость процесса стохастических аппроксимаций. Для исследования сходимости п. в. последовательности случайных величин $\{Z_n\}$ применим известный в теории устойчивости метод функций Ляпунова [46, 53]. Этот метод дает возможность в каждом конкретном случае устанавливать достаточные условия сходимости или устойчивости какого-либо процесса. Метод функций Ляпунова

основан на следующем простом факте. Если непрерывная функция $v(z)$ положительна всюду, кроме точки $z=\theta$, в которой она равна нулю, и приращение этой функции на каждом шаге (производная по времени в случае непрерывного процесса) отрицательно в силу уравнений процесса (в нашем случае в силу (34)), то процесс всегда идет к точке θ и не может остановиться, пока не достигнет точки θ. Функция $v(z)$ называется *функцией Ляпунова*. Если процесс приближения строится с помощью случайных величин, то вместо приращения (производной) отрицательным должно быть условное математическое ожидание приращения (производной).

Для сходимости п. в. процесса стохастических аппроксимаций (34) *к корню уравнения регрессии* $\psi(z)=\alpha$ *в случае независимых величин* $Y_n-\psi(Z_{n-1})$ *достаточно существования функции Ляпунова* $v(z)$, *удовлетворяющей условиям*:

1) $v(z)$ *непрерывна вместе со своими первыми и вторыми производными*;

2) $v(\theta)=0,\quad \inf\limits_{|z-\theta|>\varepsilon} v(z)>\delta(\varepsilon)>0$ *при любом* $\varepsilon>0$;

3) $\sup\limits_{|z-\theta|>\varepsilon} u(z)<-\delta_1(\varepsilon)<0$ *при любом* $\varepsilon>0$, *где*

$$u(z)=\frac{\partial^T v(z)}{\partial z}\varepsilon[\psi(z)-\alpha]; \tag{36}$$

4) $w(z,\lambda)=$

$$=M\left[(Y^T-\alpha^T)\varepsilon\frac{\partial}{\partial z}\frac{\partial^T}{\partial z}v(z+\lambda(Y-\alpha))\varepsilon(Y-\alpha)\middle|z\right]<D<\infty \tag{37}$$

при всех z, λ (*теорема Блюма*) *).

▶ Рассмотрим последовательность скалярных случайных величин $\{v(Z_n)\}$. Пользуясь формулой Тейлора, находим

$$v(Z_n)=v(Z_{n-1})+a_n\frac{\partial^T v(Z_{n-1})}{\partial z}\varepsilon(Y_n-\alpha)+$$
$$+\frac{a_n^2}{2}(Y_n^T-\alpha^T)\varepsilon\frac{\partial}{\partial z}\frac{\partial^T}{\partial z}v(Z_{n-1}+\theta_n a_n(Y_n-\alpha))\varepsilon(Y_n-\alpha), \tag{38}$$

где $\theta_n\in(0,1)$. Применив формулу полного математического ожидания (4.34) и приняв во внимание, что ус-

*) Блюм доказал достаточность этих условий для сходимости последовательности $\{Z_n\}$ п. н. [5, 6, 13]. Мы ограничимся здесь доказательством их достаточности для сходимости п. в. Функция $u(z)$, очевидно, представляет собой условное математическое ожидание дифференциала, т. е. главной части приращения функции $v(z)$, в силу уравнений процесса (34) при данном значении z.

ловное математическое ожидание величины Y_n относительно Z_{n-1} равно $\psi(Z_{n-1})$, $M[Y_n | Z_{n-1}] = \psi(Z_{n-1})$, с учетом (36) и (37) получаем

$$M \frac{\partial^T v(Z_{n-1})}{\partial z} \varepsilon (Y_n - \alpha) = M \left[\frac{\partial^T v(Z_{n-1})}{\partial z} \varepsilon M [Y_n - \alpha | Z_{n-1}] \right] =$$
$$= M \frac{\partial^T v(Z_{n-1})}{\partial z} \varepsilon [\psi(Z_{n-1}) - \alpha] = Mu(Z_{n-1}), \quad (39)$$

$$M (Y_n^T - \alpha^T) \frac{\partial}{\partial z} \frac{\partial^T}{\partial z} v (Z_{n-1} + \theta_n a_n (Y_n - \alpha)) \varepsilon (Y_n - \alpha) =$$
$$= Mw (Z_{n-1}, \theta_n a_n) < D. \quad (40)$$

На основании (38), (39) и (40)

$$Mv(Z_n) = Mv(Z_{n-1}) + a_n Mu(Z_{n-1}) + \frac{a_n^2}{2} Mw(Z_{n-1}, \theta_n a_n) =$$
$$= Mv(Z_{n-1}) - a_n \xi_n + \frac{1}{2} a_n^2 \eta_n, \quad (41)$$

где $\xi_n = -Mu(Z_{n-1})$, $\eta_n = Mw(Z_{n-1}, \theta_n a_n)$. Применяя (41) последовательно, находим

$$Mv(Z_n) = Mv(Z_0) - \sum_{k=1}^{n} a_k \xi_k + \frac{1}{2} \sum_{k=1}^{n} a_k^2 \eta_k.$$

Отсюда следует, что

$$\sum_{k=1}^{n} a_k \xi_k = Mv(Z_0) - Mv(Z_n) + \frac{1}{2} \sum_{k=1}^{n} a_k^2 \eta_k <$$
$$< Mv(Z_0) + \frac{D}{2} \sum_{k=1}^{n} a_k^2.$$

В силу условия 3) $\xi_k = -Mu(Z_{n-1}) = M|u(Z_{n-1})| \geqslant 0$. Отсюда и из второго условия (35) следует, что ряд с неотрицательными членами $\sum a_k \xi_k$ сходится. А так как ряд $\sum a_k$ в силу первого условия (35) расходится, то $\underline{\lim}\, \xi_n = 0$. Отсюда следует существование подпоследовательности $\{\xi_{n_p}\}$, сходящейся к нулю:

$$\xi_{n_p} = M \left| u \left(Z_{n_p - 1} \right) \right| \to 0.$$

Но в силу неравенства Чебышева (6.1)

$$P\left(\left| u\left(Z_{n_p}\right) \right| > \varepsilon \right) < \frac{M \left| u\left(Z_{n_p}\right) \right|}{\varepsilon}.$$

Следовательно, $P\left(\left| u\left(Z_{n_p}\right) \right| > \varepsilon \right) \to 0$ при $p \to \infty$, т. е. подпоследовательность $\{u(Z_{n_p})\}$ сходится п. в к 0. Отсюда

и из условия 3) вытекает сходимость п. в. подпоследовательности $\{Z_{n_p}\}$ к θ. Действительно, если подпоследовательность $\{Z_{n_p}\}$ не сходится п. в. к θ, то существуют такие ε, $\zeta > 0$, что $P(|Z_{n_p} - \theta| > \varepsilon) > \zeta$ при всех n_p. Но из условия 3) следует, что

$$P(|Z_{n_p} - \theta| > \varepsilon) \leqslant P(|u(Z_{n_p})| > \delta_1(\varepsilon)).$$

Таким образом, $P(|u(Z_{n_p})| > \delta_1(\varepsilon)) > \zeta$ при всех n_p, что невозможно ввиду сходимости п. в. подпоследовательности $\{u(Z_{n_p})\}$ к 0.

Осталось доказать, что и вся подпоследовательность $\{Z_n\}$ сходится п. в. к θ. Для этого заметим, что в силу непрерывности функции $v(z)$ и условия 2) подпоследовательность $\{v(Z_{n_p})\}$ сходится п. в. к 0. С другой стороны, из (38) следует, что при любых n_p и $n > n_p$

$$v(Z_n) = v(Z_{n_p}) + \sum_{k=n_p+1}^{n} a_k \frac{\partial^T v(Z_{k-1})}{\partial z} \varepsilon (Y_k - \alpha) +$$

$$+ \frac{1}{2} \sum_{k=n_p+1}^{n} a_k^2 (Y_k^T - \alpha^T) \varepsilon \frac{\partial}{\partial z} \frac{\partial^T}{\partial z} v(Z_{k-1} + \theta_k a_k (Y_k - \alpha)) \varepsilon (Y_k - \alpha).$$

Отсюда, принимая во внимание, что в силу условия 3) с вероятностью 1 имеют место неравенства

$$M\left[\frac{\partial^T v(Z_{n_p})}{\partial z} \varepsilon (Y_{n_p+1} - \alpha) \,|\, Z_{n_p}\right] =$$
$$= \frac{\partial^T v(Z_{n_p})}{\partial z} \varepsilon [\psi(Z_{n_p}) - \alpha] \leqslant 0,$$

$$M\left[\frac{\partial^T v(Z_{k-1})}{\partial z} \varepsilon (Y_k - \alpha) \,|\, Z_{n_p}\right] =$$
$$= M\left[\frac{\partial^T v(Z_{k-1})}{\partial z} \varepsilon [\psi(Z_{k-1}) - \alpha] \,|\, Z_{n_p}\right] \leqslant 0,$$

и учитывая условие 4), получаем

$$M[v(Z_n) \,|\, Z_{n_p}] \leqslant v(Z_{n_p}) + \frac{D}{2} \sum_{k=n_p+1}^{n} a_k^2.$$

Из этого неравенства и второго условия (35) следует, что при любом $\delta > 0$

$$M[v(Z_n) \,|\, Z_{n_p}] < v(Z_{n_p}) + \delta$$

при всех достаточно больших n_p и $n > n_p$. Отсюда, пользуясь неравенством Чебышева (6.1), получаем при любом

$\varepsilon > 0$

$$P(v(Z_n) > \varepsilon \,|\, v(Z_{n_p}) < \delta) < \frac{2\delta}{\varepsilon}$$

и, следовательно,

$$P(\{v(Z_n) > \varepsilon\} \cap \{v(Z_{n_p}) < \delta\}) =$$
$$= P(v(Z_{n_p}) < \delta)\, P(v(Z_n) > \varepsilon \,|\, v(Z_{n_p}) < \delta) < \frac{2\delta}{\varepsilon}.$$

Зададим теперь произвольно малое $\zeta > 0$ и выберем $\delta < \varepsilon\zeta/4$. Тогда будем иметь при всех достаточно больших n_p и $n > n_p$

$$P(\{v(Z_n) > \varepsilon\} \cap \{v(Z_{n_p}) < \delta\}) < \frac{\zeta}{2}. \qquad (42)$$

Далее, в силу сходимости п. в. к 0 подпоследовательности $\{v(Z_{n_p})\}$

$$P(\{v(Z_n) > \varepsilon\} \cap \{v(Z_{n_p}) \geqslant \delta\}) \leqslant P(v(Z_{n_p}) \geqslant \delta) < \frac{\zeta}{2} \qquad (43)$$

при всех достаточно больших n_p. Из (42) и (43) получаем

$$P(v(Z_n) > \varepsilon) = P(\{v(Z_n) > \varepsilon\} \cap \{v(Z_{n_p}) < \delta\}) +$$
$$+ P(\{v(Z_n) > \varepsilon\} \cap \{v(Z_{n_p}) \geqslant \delta\}) < \frac{\zeta}{2} + \frac{\zeta}{2} = \zeta.$$

Это доказывает сходимость п. в. к 0 всей последовательности $\{v(Z_n)\}$. В силу непрерывности $v(z)$ и условия 2) отсюда вытекает сходимость п. в. к 0 всей последовательности $\{Z_n\}$, чем и завершается доказательство. ◄

При практическом применении доказанной теоремы самым трудным, как и при исследовании устойчивости, оказывается нахождение функции Ляпунова $v(z)$.

В случае скалярных z, Y и $\psi(z)$ функцией Ляпунова может служить $v(z) = (z - \theta)^2$. Эта функция удовлетворяет первым двум условиям доказанной теоремы. При этом

$$u(z) = \frac{\partial v(z)}{\partial z}\, \varepsilon\,[\psi(z) - \alpha] = 2(z - \theta)[\psi(z) - \alpha]\,\varepsilon =$$
$$= 2\omega(z)\,\varepsilon = -2\omega(z)\,\operatorname{sgn}\omega(z) = -2\,|\omega(z)|.$$

Поэтому условие 3) удовлетворяется, если для каждого $\varepsilon > 0$ существует такое $\delta_1(\varepsilon) > 0$, что

$$\inf_{|z-\theta| > \varepsilon} |\omega(z)| > \delta_1(\varepsilon).$$

Наконец, вследствие того, что $v''(z) = 2$,

$$w(z, \lambda) = 2M[(Y - \alpha)^2 \,|\, z] = 2\{[\psi(z) - \alpha]^2 + D[Y \,|\, z]\},$$

где $D[Y|z]$ — дисперсия величины Y при данном значении z. Поэтому условие 4) удовлетворяется, если дисперсия наблюдаемой случайной величины Y ограничена сверху одним и тем же числом D, при всех z и функция $\psi(z)$ ограничена на любом конечном интервале. При этих условиях из доказанной теоремы вытекает основная теорема Роббинса и Монро, доказанная ими в работе, положившей начало развитию общей теории стохастических аппроксимаций [92].

§ 4. Рекуррентное оценивание точки экстремума регрессии

4.1. Процесс стохастических аппроксимаций. Перейдем к случаю оценивания точки экстремума скалярной функции регрессии $\psi(z)$. Предположим, что функция $\psi(z)$ непрерывна вместе со своими первыми производными $\psi'_p(z) = \partial\psi(z)/\partial z_p$ $(p = 1, \ldots, r)$ и имеет единственный экстремум — минимум — в точке $z = \theta$ *). В этом случае естественно стремиться двигаться из каждой точки z по отрицательному направлению вектора градиента, т. е. применить *метод наискорейшего спуска* [102]. При неизвестной функции $\psi(z)$ естественно за оценку производной $\psi'_p(z)$ в точке Z_{n-1} принять величину $U_{np} = (Y_{np} - Y_{n0})/c_n$, где Y_{n0} — результат наблюдения случайной величины Y при $z = Z_{n-1}$, а Y_{np} — результат наблюдения величины Y при $z = Z_{n-1} + c_n e_p$, где e_p — единичный вектор оси z_p, а c_n — некоторое число. Тогда, взяв произвольную случайную или неслучайную величину Z_0 и определив последовательность случайных величин

$$Z_n = Z_{n-1} - a_n U_n, \qquad (44)$$

где U_n — вектор с координатами $U_{np} = (Y_{np} - Y_{n0})/c_n$, можем рассчитывать на сходимость п. в. построенного процесса стохастических аппроксимаций к точке минимума θ функции $\psi(z)$.

4.2. Сходимость процесса стохастических аппроксимаций. Докажем, что *для сходимости п. в. последовательности* (44) *к точке минимума* θ *функции* $\psi(z)$ *достаточно, чтобы положительные числа* a_n *и* c_n *и функция* $\psi(z)$

---

*) Случай максимума приводится к случаю минимума изменением знака функции $\psi(z)$ и случайной величины Y.

удовлетворяли условиям

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty,$$
$$\lim_{n\to\infty} c_n = 0, \quad \sum_{n=1}^{\infty} a_n c_n < \infty, \quad \sum_{n=1}^{\infty} \left(\frac{a_n}{c_n}\right)^2 < \infty, \tag{45}$$

$$\gamma_0 < g_p(z) = \frac{\psi_p'(z)}{z_p - \theta_p} < \gamma_1 \qquad (p = 1, \ldots, r), \tag{46}$$

где γ_0 *и* γ_1 *— некоторые положительные числа, и чтобы дисперсия наблюдаемой случайной величины* Y *была при всех* z *ограничена одним и тем же числом* D, $D[Y \mid z] < D$ (обобщенная теорема Кифера и Вольфовица [36]) *).

▶ Из (44) следует

$$|Z_n - \theta|^2 = |Z_{n-1} - \theta|^2 - 2a_n (Z_{n-1}^T - \theta^T) U_n + a_n^2 |U_n|^2. \tag{47}$$

Оценим условное математическое ожидание второго слагаемого относительно Z_{n-1}. Для этого найдем сначала $M[U_n \mid Z_{n-1}]$. По определению вектора U_n, применяя формулу конечных приращений Лагранжа и учитывая, что в силу (46) $\psi_p'(z) = (z - \theta) g_p(z)$, получаем

$$M[U_{np} \mid Z_{n-1}] = \frac{M[Y_{np} \mid Z_{n-1}] - M[Y_{n0} \mid Z_{n-1}]}{c_n} =$$
$$= \frac{\psi(Z_{n-1} + c_n e_p) - \psi(Z_{n-1})}{c_n} = \psi_p'(Z_{n-1} + \vartheta_{np} c_n e_p) =$$
$$= (Z_{n-1,p} + \vartheta_{np} c_n - \theta_p) g_p(Z_{n-1} + \vartheta_{np} c_n e_p), \tag{48}$$

где $\vartheta_{np} \in (0, 1)$. Следовательно,

$$(Z_{n-1,p} - \theta_p) M[U_{np} \mid Z_{n-1}] =$$
$$= (Z_{n-1,p} - \theta_p)^2 g_p(Z_{n-1} + \vartheta_{np} c_n e_p) +$$
$$+ \vartheta_{np} c_n (Z_{n-1,p} - \theta_p) g_p(Z_{n-1} + \vartheta_{np} c_n e_p). \tag{49}$$

Отсюда и из (46) получаем

$$-(Z_{n-1,p} - \theta_p) M[U_{np} \mid Z_{n-1}] <$$
$$< -\gamma_0 (Z_{n-1,p} - \theta_p)^2 + \gamma_1 c_n |Z_{n-1,p} - \theta_p|.$$

---

*) Кифер и Вольфовиц доказали эту теорему для скалярной переменной z. Обобщение на случай векторного z при несколько иных условиях и с доказательством сходимости п. н. дал Блюм [6].

Суммируя по p и принимая во внимание, что

$$\sum_{p=1}^{r} (Z_{n-1,p}-\theta_p) M[U_{np} | Z_{n-1}] = (Z_{n-1}^T - \theta^T) M[U_n | Z_{n-1}],$$

$$\sum_{p=1}^{r} |Z_{n-1,p}-\theta_p| \leqslant \sqrt{r \sum_{p=1}^{r} (Z_{n-1,p}-\theta_p)^2} = \sqrt{r}\, |Z_{n-1}-\theta|\,^{*)},$$

приходим к неравенству

$$-(Z_{n-1}^T-\theta^T) M[U_n | Z_{n-1}] < < -\gamma_0 |Z_{n-1}-\theta|^2 + \gamma_1 \sqrt{r} c_n |Z_{n-1}-\theta|. \quad (50)$$

Кроме этой оценки, нам еще понадобится более грубая оценка для предварительного исследования. Чтобы получить такую оценку, заметим, что в силу (46) и (49) величина $-(Z_{n-1,p}-\theta_p) M[U_{np} | Z_{n-1}]$ может быть положительной только при достаточно малых отрицательных $Z_{n-1,p}-\theta_p$, а именно при $0 > Z_{n-1,p}-\theta_p > -\vartheta_{np}c_n > -c_n$. Отсюда и из (49) следует, что

$$-(Z_{n-1,p}-\theta_p) M[U_{np} | Z_{n-1}] < \gamma_1 c_n |Z_{n-1,p}-\theta_p| < \gamma_1 c_n^2.$$

Суммируя по r, получаем неравенство

$$-(Z_{n-1}^T-\theta^T) M(U_n | Z_{n-1}] < r\gamma_1 c_n^2. \quad (51)$$

Для оценки условного математического ожидания последнего слагаемого в (47) относительно Z_{n-1} воспользуемся формулой (3.29), выражающей начальный момент второго порядка через математическое ожидание и дисперсию случайной величины. На основании этой формулы, применив (48) и (46), получаем

$$M[U_{np}^2 | Z_{n-1}] = (M[U_{np} | Z_{n-1}])^2 + D[U_{np} | Z_{n-1}] < < \gamma_1^2 (Z_{n-1,p} + \vartheta_{np}c_n - \theta_p)^2 + D[U_{np} | Z_{n-1}] < < 2\gamma_1^2 [(Z_{n-1,p}-\theta_p)^2 + c_n^2] + D[U_{np} | Z_{n-1}].$$

---

*) Здесь мы воспользовались неравенством

$$\left(\sum_{p=1}^{r} x_p\right)^2 \leqslant r \sum_{p=1}^{r} x_p^2,$$

справедливым при любом r и любых действительных $x_1, \ldots, x_r$. Для доказательства достаточно найти максимум суммы $\sum u_p$ при условии $\sum u_p^2 = 1$ и показать, что он равен $\sqrt{r}$.

Но

$$D[U_{np}|Z_{n-1}] = \frac{1}{c_n^2} D[Y_{np} - Y_{n0}|Z_{n-1}] \leqslant$$
$$\leqslant \frac{1}{c_n^2}\{D[Y_{np}|Z_{n-1}] + D[Y_{n0}|Z_{n-1}] +$$
$$+ 2\sqrt{D[Y_{np}|Z_{n-1}]D[Y_{n0}|Z_{n-1}]}\} =$$
$$= \frac{1}{c_n^2}\{\sqrt{D[Y_{np}|Z_{n-1}]} + \sqrt{D[Y_{n0}|Z_{n-1}]}\}^2 < \frac{4D}{c_n^2}$$

в силу условия теоремы. Следовательно,

$$M[U_{np}^2|Z_{n-1}] < 2\gamma_1^2(Z_{n-1,\,p} - \theta_p)^2 + 2\gamma_1^2 c_n^2 + \frac{4D}{c_n^2}.$$

Отсюда суммированием по p получаем неравенство

$$M[|U_n|^2|Z_{n-1}] < 2\gamma_1^2|Z_{n-1} - \theta|^2 + 2r\gamma_1^2 c_n^2 + \frac{4rD}{c_n^2}. \quad (52)$$

Теперь мы можем доказать с. к. сходимость последовательности $\{Z_n\}$ к θ. Для этого докажем сначала ограниченность последовательности положительных чисел $\delta_n = M|Z_n - \theta|^2$. Из (47) следует, что

$$\delta_n = \delta_{n-1} - 2a_n M(Z_{n-1}^T - \theta^T)U_n + a_n^2 M|U_n|^2 =$$
$$= \delta_{n-1} - 2a_n M[(Z_{n-1}^T - \theta^T)M[U_n|Z_{n-1}]] +$$
$$+ a_n^2 M[M[|U_n|^2|Z_{n-1}]]. \quad (53)$$

Отсюда, пользуясь неравенствами (51) и (52), получаем

$$\delta_n < \delta_{n-1} + 2r\gamma_1 a_n c_n^2 + 2\gamma_1^2 a_n^2 \delta_{n-1} + 2r\gamma_1^2 a_n^2 c_n^2 +$$
$$+ 4rD\left(\frac{a_n}{c_n}\right)^2 = \delta_{n-1}(1 + 2\gamma_1^2 a_n^2) + \mu_n, \quad (54)$$

где

$$\mu_n = 2r\gamma_1 a_n c_n^2 + 2r\gamma_1^2 a_n^2 c_n^2 + 4rD\left(\frac{a_n}{c_n}\right)^2.$$

Применяя неравенство (54) последовательно, будем иметь

$$\delta_n < \delta_0 \prod_{k=1}^{n}(1 + 2\gamma_1^2 a_k^2) + \sum_{s=1}^{n-1}\mu_s \prod_{k=s+1}^{n}(1 + 2\gamma_1^2 a_k^2) + \mu_n. \quad (55)$$

Из (45) следует, что ряд $\sum \mu_s$ и бесконечное произведение $\prod (1+2\gamma_1^2 a_k^2)$ сходятся *). Таким образом, все произведения в (55) образуют неубывающие последовательности, ограниченные сверху одним и тем же числом — бесконечным произведением $\prod (1+2\gamma_1^2 a_k^2)$. Отсюда в силу сходимости ряда $\sum \mu_s$ следует ограниченность последовательности $\{\delta_n\}$, т. е. существование такого числа b, что

$$\delta_n < b^2 \text{ при любом } n.$$

А так как для любой векторной случайной величины V на основании неравенства (3.45) $M|V| = M(1\cdot|V|) \leqslant \leqslant \sqrt{M|V|^2}$, то $M|Z_n-\theta| \leqslant \sqrt{M|Z_n-\theta|^2} = \sqrt{\delta_n} < b$. Пользуясь этим неравенством и неравенствами (50) и (52), из (53) получаем

$$\delta_n < \delta_{n-1} - 2\gamma_0 a_n \delta_{n-1} + 2\gamma_1 b\sqrt{r}\, a_n c_n + 2\gamma_1^2 b^2 a_n^2 + \\ + 2r\gamma_1^2 a_n^2 c_n^2 + 4rD\left(\frac{a_n}{c_n}\right)^2 = \delta_{n-1}(1-2\gamma_0 a_n) + \nu_n, \quad (56)$$

где

$$\nu_n = \gamma_1 b\sqrt{r}\, a_n c_n + 2\gamma_1^2 b^2 a_n^2 + 2r\gamma_1^2 a_n^2 c_n^2 + 4rD\left(\frac{a_n}{c_n}\right)^2.$$

Применяя (56) последовательно, будем иметь

$$\delta_n < \delta_{n_0} \prod_{k=n_0+1}^{n} (1-2\gamma_0 a_k) + \sum_{s=n_0+1}^{n-1} \nu_s \prod_{k=s+1}^{n} (1-2\gamma_0 a_k) + \nu_n, \quad (57)$$

где n_0 — любое число, для которого $2\gamma_0 a_k < 1$ при всех $k > n_0$. Из (45) следует, что ряд $\sum \nu_s$ сходится, а про-

---

*) Так как $\ln \prod (1+x_k) = \sum \ln (1+x_k) = \sum x_k \beta(x_k)$, где $\beta(x_k) = 1 - x_k/2(1+\lambda_k x_k)$, $\lambda_k \in (0, 1)$, и $\beta(x_k) \longrightarrow 1$ при $x_k \longrightarrow 0$, то бесконечное произведение $\prod (1+x_k)$ сходится тогда и только тогда, когда сходится ряд $\sum x_k$. Если $\sum x_k = \infty$, то бесконечное произведение $\prod (1+x_k)$ называется *расходящимся к* ∞, так как в этом случае

$$\lim_{n\to\infty} \prod_{k=1}^{n} (1+x_k) = \infty.$$

Если $\sum x_k = -\infty$, то произведение $\prod (1+x_k)$ называется *расходящимся к* нулю, так как в этом случае

$$\lim_{n\to\infty} \prod_{k=1}^{n} (1+x_k) = 0.$$

изведение $\prod(1-2\gamma_0 a_k)$ расходится к нулю *). Следовательно, правая часть в (57) стремится к 0 при $n \to \infty$. Это и доказывает сходимость последовательности $\{\delta_n\}$ к нулю, т. е. с. к. сходимость последовательности случайных величин $\{Z_n\}$ к θ. Наконец, из с. к. сходимости $\{Z_n\}$ к θ следует сходимость п. в., что и доказывает теорему. ◀

4.3. Метод случайного поиска. Мы видим, что для оценки экстремума функции $\psi(z)$ r-мерного вектора z необходимо в каждой точке Z_{n-1} наблюдать случайную величину Y $r+1$ раз при $z=Z_{n-1}$, $Z_{n-1}+c_n e_1$, ..., $Z_{n-1}+c_n e_r$. При $r=1$ величину Y приходится наблюдать 2 раза. При этом вместо наблюдения ее при $z=Z_{n-1}$ и $z=Z_{n-1}+c_n$ можно наблюдать ее при $z=Z_{n-1}+c_n$ и $z=Z_{n-1}-c_n$ и за оценку производной функции $\psi(z)$ в точке Z_{n-1} принять отношение разности полученных значений Y к $2c_n$. Именно этот случай первоначально рассматривали Кифер и Вольфовиц в [36].

При большой размерности r вектора z применение метода стохастических аппроксимаций для нахождения экстремума функции $\psi(z)$ оказывается очень трудоемким из-за необходимости делать $r+1$ наблюдений величины Y в каждой точке. В таких случаях более выгодным часто оказывается метод *случайного поиска*.

Метод случайного поиска основан на том, что из каждой точки Z_{n-1} делается шаг, представляющий собой случайный вектор V_n. В полученной точке $Z_n=Z_{n-1}+V_n$ производится одно наблюдение случайной величины Y. В соответствии с результатом этого наблюдения Y_n формируется распределение следующего шага V_{n+1}. Если $Y_n < Y_{n-1}$, то распределение шага V_{n+1} выбирается так, чтобы вероятность шага V_{n+1}, близкого по направлению к V_n (т. е., вероятность сделать следующий шаг в направлении, близком к тому, в котором была получена удача), была достаточно велика. При $Y_n > Y_{n-1}$ распределение шага V_{n+1} выбирается так, чтобы вероятность шага V_{n+1}, близкого по направлению к V_n (т. е. вероятность сделать следующий шаг в направлении, близком к тому, в котором была получена неудача), была малой. В остальном распределения пробных шагов можно брать в значительной мере произвольными. При удачном выборе этих распределений

---

*) См. сноску на стр. 295 с учетом того, что $-2\gamma_0 \sum a_k = -\infty$.

метод случайного поиска может дать значительный выигрыш по сравнению с методом стохастических аппроксимаций в случае большой размерности r вектора z.

Поскольку распределение каждого шага формируется с учетом результатов предыдущих шагов, алгоритм случайного поиска получается, по существу, обучающимся. Такой алгоритм накапливает опыт, учится на предыдущих удачах и неудачах *).

---

*) Более подробное изложение метода стохастических аппроксимаций и его применений читатель может найти в [13, 102]. Приложения метода стохастических аппроксимаций к различным задачам управления даны в [47, 106, 121, 122]. Читателей, интересующихся методом случайного поиска и его многочисленными приложениями, мы отсылаем к книге Л. А. Растригина [87], который внес существенный вклад в разработку метода случайного поиска.

Глава 8

ОЦЕНИВАНИЕ РАСПРЕДЕЛЕНИЙ

§ 1. Оценки плотности и функции распределения

1.1. Параметрическое и непараметрическое оценивание распределений. Методы оценивания числовых характеристик случайных величин и параметров распределений, изложенные в главах 6 и 7, дают возможность находить по результатам опытов (оценивать) распределения случайных величин, представляющие собой известные функции, зависящие от конечного числа неизвестных скалярных параметров (т. е. от неизвестного конечномерного параметра). Такое оценивание неизвестных распределений называется *параметрическим*. Непосредственное оценивание распределений без предположения, что они являются известными с точностью до конечномерного параметра функциями, называется *непараметрическим*. Мы изложим теперь основные методы непараметрического оценивания распределений случайных величин.

1.2. Оценивание плотности по гистограмме. Из нашего подхода к определению плотности в п. 2.2.1 следует, что подходящей оценкой плотности случайной величины может служить относительная плотность экспериментальных точек. При этом, чем больше число опытов, тем мельче можно взять интервалы разбиения области возможных значений случайной величины и тем точнее гистограмма будет аппроксимировать плотность. Практически интервалы рекомендуется выбирать так, чтобы в каждом из них было не меньше 10 экспериментальных точек *).

Пусть $f(x)$ — неизвестная плотность случайной величины X. Предположим, что область возможных значений величины X разбита на r интервалов (многомерных интервалов — прямоугольников — в случае векторной вели-

*) Для построения гистограмм на ЭВМ в [94] служат программы HIST, TAB1 (одномерные гистограммы) и AB2 (двумерные гистограммы).

чины X) $\Delta x_1, \ldots, \Delta x_r$ *). Пусть $Z_1, \ldots, Z_r$ — случайные числа попаданий величины X в интервалы $\Delta x_1, \ldots, \Delta x_r$ при n опытах, $Z_1 + \ldots + Z_r = n$. Тогда частоты попадания в эти интервалы будут равны $\hat{P}_\nu = Z_\nu / n$ ($\nu = 1, \ldots, r$). Вероятности попадания в эти интервалы определяются формулой

$$p_\nu = \int_{\Delta x_\nu} f(x)\,dx \qquad (\nu = 1, \ldots, r).$$

Если принять частоты $\hat{P}_1, \ldots, \hat{P}_r$ за оценки вероятностей $p_1, \ldots, p_r$, то значения относительной плотности экспериментальных точек $\hat{F}_\nu = \hat{P}_\nu / \Delta x_\nu$ в соответствующих интервалах Δx_ν ($\nu = 1, \ldots, r$) будут оценками величин

$$f_\nu = \frac{p_\nu}{\Delta x_\nu} = \frac{1}{\Delta x_\nu} \int_{\Delta x_\nu} f(x)\,dx \qquad (\nu = 1, \ldots, r)\ **).$$

Если $f(x)$ непрерывна в каждом интервале Δx_ν, то эти величины представляют собой значения $f(x)$ в некоторых средних точках соответствующих интервалов. Таким образом, значения относительной плотности экспериментальных точек (ординаты гистограммы) представляют собой оценки плотности вероятности в неизвестных средних точках соответствующих интервалов.

1.3. Доверительные области для плотности. Для нахождения доверительной области для вектора $f = [f_1 \ldots f_r]^T$ в r-мерном пространстве воспользуемся третьим методом п. 6.2.4. Чтобы найти соответствующую функцию φ, будем оценивать отклонение вектора частот $\hat{P} = [\hat{P}_1 \ldots \hat{P}_r]^T$ от вектора вероятностей $p = [p_1 \ldots p_r]^T$ взвешенной суммой квадратов ошибок

$$\sum_{\nu=1}^{r} c_\nu (\hat{P}_\nu - p_\nu)^2,$$

где $c_1, \ldots, c_r$ — некоторые положительные числа. Эти числа целесообразно выбрать так, чтобы меньшие значения вероятности оценивать с меньшими абсолютными

---

*) Если областью возможных значений величины X служит все пространство, то некоторые из интервалов $\Delta x_1, \ldots, \Delta x_r$ могут быть бесконечными. При оценивании плотности все интервалы, содержащие экспериментальные точки, берутся конечными и только интервалы, не содержащие экспериментальных точек могут быть бесконечными.

**) Для простоты мы обозначаем через Δx_ν не только интервал, но и его длину (объем в случае векторной X).

ошибками. Иными словами, ошибкам оценивания малых вероятностей p_ν следует приписать большие веса c_ν. К. Пирсон показал, что если взять $c_\nu = n/p_\nu$ $(\nu = 1, \ldots, r)$, то в пределе при неограниченном возрастании числа опытов n распределение величины

$$\mathrm{H} = n \sum_{\nu=1}^{r} \frac{1}{p_\nu} (\hat{P}_\nu - p_\nu)^2 \tag{1}$$

стремится к χ^2-распределению с $r-1$ степенями свободы [72, 73] (пример 5.48). Так как это распределение не зависит от неизвестных параметров $p_1, \ldots, p_r$, то величину (1), называемую обычно χ^2, можно принять за функцию $\varphi(\hat{P}, p)$ п. 6.2.4. Тогда получим доверительную область для вектора $p = [p_1 \ldots p_r]^T$, определяемую соотношениями

$$\sum_{\nu=1}^{r} \frac{1}{p_\nu} (p_\nu - \hat{P}_\nu)^2 < \frac{\varepsilon_\alpha}{n}, \quad \sum_{\nu=1}^{r} p_\nu = 1, \tag{2}$$

где ε_α определяется по заданному коэффициенту доверия α из уравнения

$$P(\varphi(\hat{P}, p) < \varepsilon_\alpha) = P(\mathrm{H} < \varepsilon_\alpha) = \int_0^{\varepsilon_\alpha} p_{r-1}(\eta)\, d\eta = \alpha. \tag{3}$$

Практически величину ε_α определяют по заданным $k = r-1$ и α по таблице 5. При этом следует учитывать и бесконечные интервалы Δx_ν, частоты попадания в которые равны нулю.

Для нахождения доверительной области для вектора средних плотностей $f = [f_1 \ldots f_r]^T$ следует прежде всего заменить бесконечные интервалы Δx_ν достаточно большими конечными так, чтобы вероятность непопадания величины X ни в один из интервалов $\Delta x_1, \ldots, \Delta x_r$ была пренебрежимо малой. При этом для более точного интервального оценивания средних плотностей в бесконечных интервалах с нулевыми частотами попадания можно разбить их на несколько меньших интервалов (с нулевыми частотами). Тогда доверительная область для f определителя соотношениями

$$\sum_{\nu=1}^{r} \frac{1}{f_\nu} (f_\nu - \hat{F}_\nu)^2 \Delta x_\nu < \frac{\varepsilon_\alpha}{n}, \quad \sum_{\nu=1}^{r} f_\nu \Delta x_\nu = 1. \tag{4}$$

Пример 1. Полученные в результате 100 опытов значения случайной величины X распределились по 10 интервалам длины 80 единиц каждый, как показано в таблице.

ν	1	2	3	4	5	6	7	8	9	10
z_ν	2	6	8	16	19	17	14	8	7	3
$\bar{x}_\nu$	—345	—265	—185	—105	—25	55	135	215	305	385
$\hat{f}_\nu \cdot 10^4$	2,5	7,5	10,0	20,0	23,8	21,2	17,5	10,0	8,8	3,7

В третьей строке таблицы показаны средние точки интервалов, а в четвертой — относительные плотности экспериментальных точек. Найти оценку плотности и доверительную область для нее, соответствующую коэффициенту доверия $\alpha = 0{,}95$.

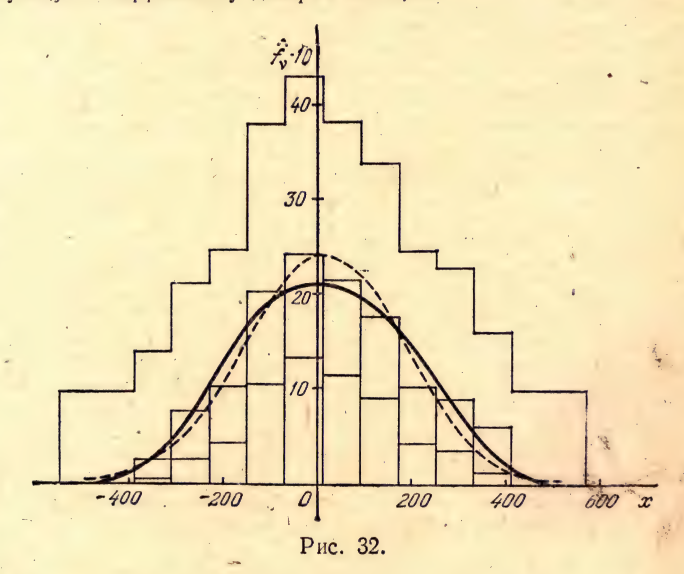


Рис. 32.

Гистограмма для данного случая дана на рис. 32. Чтобы найти доверительную область, заметим, что в данном случае, считая два бесконечных интервала $(-\infty, -385)$ и $(415, \infty)$, имеем $r = 12$, $\Delta x_1 = \ldots = \Delta x_{10} = 80$. Заменив бесконечные интервалы приближенно конечными $(-545, -385)$ и $(415, 575)$, будем иметь $\Delta x_{11} = \Delta x_{12} = 160$. По таблице 5 для $k = r - 1 = 11$ и $\alpha = 0{,}95$ находим $\varepsilon_\alpha = 19{,}68$. Полученная в результате опытов реализация доверительной области

для $f=[f_1 \dots f_{12}]^T$ определяется соотношениями

$$\frac{1}{f_1}(f_1-2{,}5\cdot10^{-4})^2+\frac{1}{f_2}(f_2-7{,}5\cdot10^{-4})^2+\frac{1}{f_3}(f_3-10{,}0\cdot10^{-4})^2+$$
$$+\frac{1}{f_4}(f_4-20{,}0\cdot10^{-4})^2+\frac{1}{f_5}(f_5-23{,}8\cdot10^{-4})^2+\frac{1}{f_6}(f_6-21{,}2\cdot10^{-4})^2+$$
$$+\frac{1}{f_7}(f_7-17{,}5\cdot10^{-4})^2+\frac{1}{f_8}(f_8-10{,}0\cdot10^{-4})^2+\frac{1}{f_9}(f_9-8{,}8\cdot10^{-4})^2+$$
$$+\frac{1}{f_{10}}(f_{10}-3{,}7\cdot10^{-4})^2+2f_{11}+2f_{12}<0{,}00246,$$
$$f_1+f_2+\dots+2f_{11}+2f_{12}=0{,}0125.$$

Эту 11-мерную область в 12-мерном пространстве невозможно представить наглядно. Чтобы получить наглядную доверительную область, найдем совместные доверительные интервалы для всех величин $f_1, \dots, f_{12}$, соответствующие коэффициенту доверия не меньше чем $\alpha=0{,}95$. Для этого следует найти доверительные интервалы для каждой из величин $f_1, \dots, f_{12}$, соответствующие коэффициенту доверия $\alpha_1=1-(1-\alpha/12)=1-0{,}05/12\approx0{,}9959$ (п. 6.2.4). Так как число опытов в данном случае достаточно велико для того, чтобы распределение частоты можно было считать нормальным, применим формулу (6.11). Для этого по таблице 1 находим (для $\alpha=0{,}9959$) $\varepsilon_\alpha=2{,}86$. Подставив это значение и значения частот в (6.11), находим совместные доверительные интервалы для $f_1, \dots, f_{12}$. Границы этих интервалов (a_ν, b_ν) $(\nu=1, \dots, 12)$ приведены в таблице.

f_ν	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}, f_{12}
$a_\nu\cdot10^4$	0,4	2,5	3,9	10,2	12,9	11,1	8,8	3,9	3,3	1,0	0
$b_\nu\cdot10^4$	13,6	20,8	24,2	37,5	40,5	37,7	33,3	24,2	22,5	15,5	9,5

На рис. 32 показана доверительная область для кривой распределения, образованная этими доверительными интервалами. Согласно сказанному в п. 6.2.4 эта доверительная область содержит неизвестную кривую распределения с вероятностью, не меньшей чем 0,95. Из рис. 32 видно, что полученные экспериментальные данные могут соответствовать большому разнообразию кривых распределения. Сплошной линией на рис. 32 показана кривая распределения, подобранная для этих экспериментальных данных в примере 5, а штриховой линией — нормальная кривая распределения.

Изложенным методом определения доверительных областей для плотности можно пользоваться только в том случае, когда число опытов n достаточно велико для того, чтобы распределение величины H мало отличалось от χ^2-распределения. Практически этим методом можно пользоваться при $n\geqslant100$. Это условие обычно всегда выполняется, так как хорошие оценки распределения

случайной величины можно получить только по результатам большого числа опытов, порядка нескольких сотен для скалярной и тысяч для векторной случайной величины. Если же возникает необходимость оценить плотность по малому числу опытов, то для нахождения доверительной области для нее следует пользоваться точным полиномиальным распределением частот $\hat{P}_1, \ldots, \hat{P}_r$ и и вторым методом п. 6.2.4, так же как это было сделано для оценки вероятности в п.6.3.2.

1.4. Оценка функции распределения. Для получения оценки функции распределения $F(x)$ естественно заменить в ее определении (2.29) вероятности событий $X < x$ их частотами. Пусть $M(x)$ — случайное число экспериментальных значений случайной величины X при n опытах, удовлетворяющих неравенству $X_k < x$. Тогда оценка функции распределения величины X определится формулой

$$\hat{F}(x) = M\ x)/n.$$

Пользуясь единичной ступенчатой функцией, можно представить эту формулу в виде *)

$$\hat{F}(x) = \frac{1}{n} \sum_{k=1}^{n} 1(x - X_k). \tag{5}$$

1.5. Доверительные области для функции распределения. Задача нахождения доверительной области для функции распределения очень сложна и для векторных величин пока еще не решена. Для скалярных случайных величин эта задача упрощается вследствие того, что для любой непрерывной функции распределения $F(x)$ распределение случайной величины $S = \max|\hat{F}(x) - F(x)|$ не зависит от $F(x)$. Это следует из результата п. 5.2.3, согласно которому для любой непрерывной $F(x)$ случайная величина $Y = F(X)$ в случае скалярной величины X распределена равномерно на интервале (0, 1).

Пусть $G_n(s)$ — функция распределения величины S при n опытах. Значения $G_n(s)$ при $s = k/n$, $k = 1, \ldots$ $\ldots, n-1$, определяются формулой [58] (см. также [103])

$$G_n\left(\frac{k}{n}\right) = \frac{n!}{n^n} u_{kn}(k, n) \qquad (k = 1, \ldots, n-1), \tag{6}$$

*) Единичную ступенчатую функцию векторного аргумента можно считать равной произведению единичных ступенчатых функций всех координат вектора-аргумента.

где функция $u_{kn}(h, m)$ $(h=1, \ldots, k;\ m=0, 1, \ldots, n)$ определяется рекуррентной формулой

$$u_{kn}(h, m+1) = \sum_{i=1}^{h+1} \frac{u_{kn}(i, m)}{(h+1-i)!}$$

и условиями

$$u_{kn}(h, m) = 0 \quad \text{при} \quad h \geqslant m+k,$$
$$u_{kn}(h, 0) = 0 \quad (h = 1, \ldots, k-1), \quad u_{kn}(k, 0) = 1.$$

Пользуясь формулой (6), Масси составил таблицы значений $G_n(k/n)$ функций $G_n(s)$, а затем интерполированием получил таблицы значений функций $G_n(s)$ в зависимости от величины $u = s\sqrt{n}$ при разных n или, что то же, таблицы значений функций распределения $K_n(u)$ случайной величины $U = S\sqrt{n} = \sqrt{n}\max|\hat{F}(x) - F(x)|$ при разных значениях числа опытов n. По этим таблицам можно находить точные доверительные полосы для оцениваемых функций распределения скалярных случайных величин. А именно доверительная область для $F(x)$, соответствующая коэффициенту доверия α, определяется неравенством

$$|F(x) - \hat{F}(x)| < u_\alpha/\sqrt{n} \quad \text{для всех } x, \tag{7}$$

где u_α находится путем решения уравнения $K_n(u_\alpha) = \alpha$ с помощью таблиц функций $K_n(u_\alpha)$. Неравенство (7) определяет случайную полосу, ограниченную снизу и сверху двумя ступенчатыми линиями, внутри которой с вероятностью α полностью содержится график неизвестной функции распределения $F(x)$. Вследствие этого в среднем только в $100(1-\alpha)\%$ случаев реализация доверительной области, определяемой неравенством (7), не будет полностью содержать график функции распределения $F(x)$ *).

Заметим, что еще задолго до работы Масси А. Н. Колмогоров нашел предельное (асимптотическое) распределение величины U при $n \to \infty$ [41]. Он доказал, что при

*) Заметим, что, пользуясь доверительным интервалом п. 6.3.2 для вероятности, можно получить значительно меньший доверительный интервал для значения $F(x)$ при каждом отдельно взятом x. Однако такие доверительные интервалы не имеют никакого практического значения, так как при оценивании функции распределения необходимо найти вероятные границы для нее сразу при всех x.

любом u

$$K(u) = \lim_{n \to \infty} K_n(u) = 1 - 2 \sum_{\nu=1}^{\infty} (-1)^{\nu+1} e^{-2\nu^2 u^2}. \qquad (8)$$

Поэтому при достаточно большом числе опытов n величину u_α в (7) можно определять из уравнения $K(u_\alpha) = \alpha$ по таблице функции Колмогорова $K(u)$ (приложение 3, табл. 7).

Пример 2. В условиях предыдущего примера найти доверительную полосу для $F(x)$, соответствующую коэффициенту доверия $\alpha = 0{,}95$. Так как в данном случае $n = 100$, то, определив по таблице 7 $u_\alpha = 1{,}362$, получим $u_\alpha/10 = 0{,}1362$. Сдвинув на эту величину вверх и вниз полученный по результатам опытов график эмпирической функции распределения и имея в виду, что $0 \leqslant F(x) \leqslant 1$, получим доверительную полосу для $F(x)$, представленную на рис. 33. График неизвестной $F(x)$ с вероятностью $\alpha = 0{,}95$ полностью заключен внутри этой полосы.

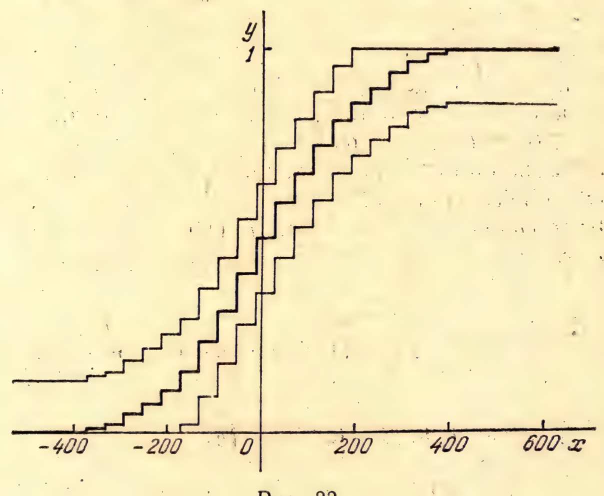


Рис. 33.

1.6. Другие оценки плотности. Дифференцируя формулу (5) и учитывая, что производная единичной ступенчатой функции представляет собой δ-функцию, получаем оценку плотности величины X в виде

$$\hat{f}(x) = \frac{1}{n} \sum_{k=1}^{n} \delta(x - X_k). \qquad (9)$$

Ясно, что в случае непрерывной величины X эта оценка не может быть близкой к $f(x)$ ни при каком значении x ни при каком числе опытов n. Эта оценка несостоятельна. Чтобы исправить ее и сделать состоятельной, можно заменить δ-функцию некоторой непрерывной плотностью $\psi_n(t)$. Тогда оценка плотности определится формулой

$$\hat{f}(x) = \frac{1}{n}\sum_{k=1}^{n} \psi_n(x - X_k). \tag{10}$$

Плотность ψ_n естественно выбирать по-разному для различных n. Чем больше n, тем ближе должна быть $\psi_n(t)$ к δ-функции $\delta(t)$. С другой стороны, в выражении (10) функции $\hat{f}(x)$ при $n \to \infty$ не должны возникать δ-функции. Это будет обеспечено, если $\psi_n(t)$ будет стремиться к $\delta(t)$ достаточно «медленно» для того, чтобы $\psi_n(t)/n \to 0$ для всех t при $n \to \infty$. После этих наводящих рассуждений можно перейти к строгому доказательству следующей теоремы.

Для состоятельности оценки (10) *плотности* $f(x)$ *в каждой точке непрерывности* $f(x)$ *достаточно выполнения условий*:

$$1) \qquad \int_{-\infty}^{\infty} \psi_n(x - z) f(z)\, dz \to f(x) \quad \textit{при}\ n \to \infty \tag{11}$$

в каждой точке непрерывности $f(x)$;

$$2) \qquad \frac{1}{n}\sup \psi_n(t) \to 0 \quad \textit{при}\ n \to \infty. \tag{12}$$

► Для доказательства найдем математическое ожидание и второй момент оценки $\hat{f}(x)$. Так как все экспериментальные значения $X_1, \ldots, X_n$ случайной величины X имеют плотность $f(x)$, то

$$M\hat{f}(x) = \frac{1}{n}\sum_{\nu=1}^{n} M\psi_n(x - X_\nu) = \int_{-\infty}^{\infty} \psi_n(x - z) f(z)\, dz.$$

В силу условия (11) эта величина стремится к $f(x)$ при $n \to \infty$ в каждой точке непрерывности $f(x)$, $M\hat{f}(x) \to f(x)$.

Далее, вследствие независимости величин $X_1, \ldots, X_n$

$$M\hat{f}^2(x) = \frac{1}{n^2}\sum_{\nu=1}^{n} M\psi_n^2(x - X_\nu) +$$
$$+ \frac{1}{n^2}\sum_{\mu\neq\nu} M\psi_n(x - X_\nu)\, M\psi_n(x - X_\mu) =$$
$$= \frac{1}{n}\int_{-\infty}^{\infty} \psi_n^2(x-z)\, f(z)\, dz + \frac{n-1}{n}\left[\int_{-\infty}^{\infty} \psi_n(x-z)\, f(z)\, dz\right]^2.$$

Но

$$\frac{1}{n}\int_{-\infty}^{\infty} \psi_n^2(x-z)\, f(z)\, dz \leqslant \frac{1}{n}\sup_t \psi(t) \int_{-\infty}^{\infty} \psi_n(x-z)\, f(z) dz.$$

В силу (12) правая часть этого неравенства стремится к нулю при $n \to \infty$. Поэтому $M\hat{f}^2(x)$ в силу (11) и (12) стремится к $f^2(x)$ при $n \to \infty$ в каждой точке непрерывности $f(x)$, $M\hat{f}^2(x) \to f^2(x)$. Следовательно,

$$M[\hat{f}(x) - f(x)]^2 = M\hat{f}^2(x) - 2f(x)\, M\hat{f}(x) + f^2(x) \to 0 \quad \text{при} \quad n \to \infty$$

в каждой точке непрерывности $f(x)$. Таким образом, при выполнении условий (11) и (12) оценка $\hat{f}(x)$ с. к. сходится к $f(x)$ в каждой точке непрерывности $f(x)$. Но из с. к. сходимости по теореме п. 6.1.3 вытекает сходимость п. в., т. е. состоятельность $\hat{f}(x)$ в каждой точке непрерывности $f(x)$.

Заметим, что если $f(x)$ непрерывна при всех x и сходимость в (11) равномерна относительно x, то $\hat{f}(x)$ равномерно с. к. и п. в. сходится к $f(x)$. ◄

Условия (11) и (12) выполняются, и, следовательно, оценка (10) состоятельна, в частности, если $\psi_n(t) = |c_n|^m \psi(c_n t)$, где $\psi(s)$ — произвольная ограниченная плотность, а c_n — числа, удовлетворяющие условиям $|c_n| \to \infty$, $c_n^m/n \to 0$ при $n \to \infty$, где m — размерность случайного вектора X. При этом для равномерной сходимости $\hat{f}(x)$ к $f(x)$ достаточна равномерная непрерывность $f(x)$ [23, 62, 70, 93, 125].

Доказанная теорема не решает вопрос о выборе плотностей $\psi_n(t)$. Она оставляет для этого широкий простор. Это дает возможность ставить задачи оптимизации оценок вида (10) при данном n путем выбора $\psi_n(t)$. Однако в

общем случае решение задачи оптимизации зависит от неизвестной плотности $f(x)$ и вследствие этого не имеет смысла с практической точки зрения.

Заметим еще, что оценки вида (10) можно получить и для производных плотности $f(x)$ [23, 125].

§ 2. Приближенное представление распределений

2.1. Система кривых Пирсона. В приложениях часто требуется приближенно представить плотность или функцию распределения случайной величины подходящим аналитическим выражением. В частности, приходится находить подходящие аналитические выражения для распределений, оцениваемых по результатам опытов. Это можно сделать многими способами. Так, например, формула (10) непосредственно дает оценку плотности в виде линейной комбинации определенных функций. Однако этот единственный способ получения непараметрической оценки плотности в аналитической форме обладает тем недостатком, что число слагаемых в (10), равное числу опытов n, как правило, очень велико. Всякое же упрощение аналитического представления плотности неизбежно ведет к параметризации оценки, т. е. к замене непараметрической оценки параметрической.

Для приближенного аналитического представления плотностей скалярных случайных величин широко применяется система кривых распределения К. Пирсона [71, 74]. Он заметил, что большая часть плотностей $y=f(x)$, встречающихся в задачах практики, удовлетворяет дифференциальному уравнению вида

$$y'=\frac{x+a}{b_0+b_1x+b_2x^2}\,y,$$

где a, b_0, b_1 и b_2 — некоторые постоянные. В частности, этому уравнению удовлетворяют нормальное, показательное, стьюдентово распределения, β-, γ-, χ-, χ^2-распределения. Изменяя параметры a, b_0, b_1 и b_2, можно получить большое разнообразие кривых распределения.

Другая система кривых для приближенного представления плотностей скалярных случайных величин, основанная на анализе причин отклонения от нормального распределения в производстве, предложена Бородачевым [10].

2.2. Ортогональные разложения плотностей. Наиболее общим способом приближенного представления распределений скалярных и векторных случайных величин является представление их в форме линейных комбинаций некоторых заданных функций, в частности в форме конечных отрезков разложений в различные ряды. Мы изложим здесь способ разложения плотностей по ортогональным полиномам и общий способ приближения их линейными комбинациями заданных функций.

Пусть $f(x)$ — плотность, которую требуется представить аналитическим выражением, $f_0(x)$ — некоторая «эталонная» плотность, $p_n(x)$ $(n = 0, 1, \ldots)$ — система полиномов, ортонормированных относительно распределения $f_0(x)$, т. е. удовлетворяющих условию

$$\int_{-\infty}^{\infty} p_\nu(x)\, p_\mu(x)\, f_0(x)\, dx = \delta_{\nu\mu}, \tag{13}$$

где, как всегда, $\delta_{\nu\mu} = 0$ при $\mu \neq \nu$, $\delta_{\nu\nu} = 1$. Система полиномов $p_n(x)$, удовлетворяющих условию (13), может быть найдена путем применения известного способа ортонормирования функций ко всем степенным функциям координат вектора x [42]. При этом всегда $p_0(x) = 1$ вследствие свойства (2.11) плотности, а $p_n(x)$ в случае скалярного x представляет собой полином степени n:

$$p_0(x) = 1, \quad p_n(x) = \sum_{k=0}^{n} a_{nk} x^k \qquad (n = 1, 2, \ldots). \tag{14}$$

Будем искать формальное разложение плотности $f(x)$ в ряд вида

$$f(x) = f_0(x) \sum_{m=0}^{\infty} c_m p_m(x). \tag{15}$$

Для определения коэффициентов c_n умножим (15) на $p_n(x)$ и проинтегрируем результат от $-\infty$ до ∞. Тогда, принимая во внимание (13), получим

$$c_n = \int_{-\infty}^{\infty} p_n(x)\, f(x)\, dx \qquad (n = 0, 1, 2, \ldots). \tag{16}$$

Эту формулу можно символически записать в виде

$$c_0 = 1, \ c_n = p_n(\alpha), \tag{17}$$

где $p_n(\alpha)$ представляет собой результат замены в выражении полинома $p_n(x)$ всех одночленов вида $x_1^{k_1} \dots x_n^{k_n}$ соответствующими моментами $\alpha_{k_1, \dots, k_n}$ случайной величины X, имеющей плотность $f(x)$. В частности, в случае скалярного x на основании (14)

$$c_0 = 1, \quad c_n = a_{n0} + \sum_{k=2}^{n} a_{nk} \alpha_k \qquad (n = 1, 2, \dots), \quad (18)$$

где α_k — момент k-го порядка случайной величины X с плотностью $f(x)$. Таким образом, все коэффициенты ряда (15) легко определяются, если известны моменты случайной величины X. При этом, конечно, необходимо существование конечных моментов всех порядков.

Ряд (15) при некоторых условиях сходится к $f(x)$ и, следовательно, может служить для аналитического представления $f(x)$ с любой степенью точности. Однако сходимость или расходимость ряда (15) не имеет практического значения, важно лишь, чтобы плотность $f(x)$ могла быть с достаточной точностью представлена с помощью небольшого числа (обычно двух-трех) членов ряда. Кроме того, обычно бывают более или менее точно известными только несколько первых моментов случайной величины, а относительно моментов высших порядков мы не знаем даже, существуют ли они. Поэтому разложением (15) пользуются, не интересуясь вопросом о его сходимости. Практика показывает, что большую часть встречающихся в приложениях распределений удается с достаточной точностью представить отрезком разложения (15), так же как и кривыми Пирсона.

Перед тем как представлять плотность величины X разложением (15), удобно предварительно нормировать ее, т.е. заменить случайной величиной $X' = K_x^{-1/2}(X - m_x)$*). Эта случайная величина имеет математическое ожидание, равное нулю, единичную ковариационную матрицу и то же распределение, что и величина X (пример 5.28). Следовательно, для нормированной скалярной случайной величины $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_k = \mu_k$ $(k = 2, 3, \dots)$ и форму-

---

*) Если $K_x = A \Lambda A^T$, где Λ — диагональная матрица с элементами $\lambda_1, \dots, \lambda_m$, а A — ортогональная матрица, то по определению $K_x^\nu = A \Lambda^\nu A^T$ при любом ν, где Λ^ν — диагональная матрица с элементами $\lambda_1^\nu, \dots, \lambda_m^\nu$.

лы (18) имеют вид

$$c_0=1,\ c_1=a_{10},\ c_2=a_{20}+a_{22}, \tag{19}$$

$$c_n=a_{n0}+a_{n2}+\sum_{k=3}^{n} a_{nk}\mu_k \qquad (n=3, 4, \ldots).$$

2.3. Разложение плотности по полиномам Эрмита. Разложение (15) можно строить на основе различных «эталонных» функций $f_0(x)$ как для непрерывных, так и для дискретных и для непрерывно-дискретных случайных величин. Выбор «эталонной» плотности $f_0(x)$ в значительной мере определяется характером функции $f(x)$. От удачного выбора $f_0(x)$ зависит простота и точность представления плотности $f(x)$ отрезком разложения (15).

Для непрерывных величин широкое распространение получило разложение (15), основанное на нормальном распределении. В этом случае для нормированной скалярной величины в силу (3.90)

$$f_0(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

При этом полиномы p_n определяются формулой

$$p_n(x)=\frac{1}{\sqrt{n!}}H_n(x) \qquad (n=0, 1, 2, \ldots), \tag{20}$$

где $H_n(x)$ — полиномы Эрмита:

$$\begin{gathered} H_0(x)=1, \quad H_1(x)=x, \quad H_2(x)=x^2-1, \\ H_3(x)=x^3-3x, \quad H_4(x)=x^4-6x^2+3, \end{gathered} \tag{21}$$

и вообще

$$H_n(x)=x^n+\sum_{m=1}^{[n/2]}(-1)^m(2m-1)!!\,C_n^{2m}x^{n-2m}. \tag{22}$$

В этом случае формулы (19) дают для первых семи коэффициентов c_n выражения

$$\begin{gathered} c_0=1, \quad c_1=c_2=0, \quad c_3=\frac{\mu_3}{\sqrt{3!}}, \quad c_4=\frac{\mu_4-3}{\sqrt{4!}}, \\ c_5=\frac{\mu_5-10\mu_3}{\sqrt{5!}}, \quad c_6=\frac{\mu_6-15\mu_4+30}{\sqrt{6!}} \end{gathered} \tag{23}$$

и ряд (15) принимает вид

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\left[1+\sum_{n=3}^{\infty}\frac{c_nH_n(x)}{\sqrt{n!}}\right]. \tag{24}$$

Вследствие того, что

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} H_n(x) = (-1)^n f_0^{(n)}(x) = (-1)^n \Phi^{(n+1)}(x),$$

где $\Phi(x)$ — функция Лапласа, определяемая формулой (3.91), формулу (24) можно также представить в виде

$$f(x) = \Phi'(x) + \sum_{n=3}^{\infty} \frac{(-1)^n c_n}{\sqrt{n!}} \Phi^{(n+1)}(x). \qquad (25)$$

Это разложение обычно называется *рядом Грама — Шарлье типа А*.

Так как для любой скалярной случайной величины X момент порядка k соответствующей нормированной величины равен μ_k/σ^k, где μ_k — центральный момент порядка k величины X, а σ — ее среднее квадратическое отклонение, то для произвольной случайной величины X, имеющей среднее квадратическое отклонение σ и математическое ожидание, отличное от нуля, величины μ_k в формулах (19) и (23) должны быть заменены соответствующими величинами μ_k/σ^k. В результате формулы (23) и (25) примут вид

$$c_0 = 1, \quad c_1 = c_2 = 0, \quad c_3 = \frac{1}{\sqrt{3!}} \frac{\mu_3}{\sigma^3}, \quad c_4 = \frac{1}{\sqrt{4!}} \left(\frac{\mu_4}{\sigma^4} - 3 \right),$$

$$c_5 = \frac{1}{\sqrt{5!}} \left(\frac{\mu_5}{\sigma^5} - 10 \frac{\mu_3}{\sigma^3} \right), \quad c_6 = \frac{1}{\sqrt{6!}} \left(\frac{\mu_6}{\sigma^6} - 15 \frac{\mu_4}{\sigma^4} + 30 \right), \qquad (26)$$

$$f(x) = \frac{1}{\sigma} \left[\Phi'\left(\frac{x - m_x}{\sigma} \right) + \sum_{n=3}^{\infty} \frac{(-1)^n c_n}{\sqrt{n!}} \Phi^{(n+1)} \left(\frac{x - m_x}{\sigma} \right) \right]. \qquad (27)$$

Сравнив формулы для c_3 и c_4 с (3.83) и (3.84), видим, что коэффициенты c_3 и c_4 пропорциональны соответственно асимметрии γ_1 и эксцессу γ_2, $c_3 = \gamma_1/\sqrt{3!}$, $c_4 = \gamma_2/\sqrt{4!}$.

При применении конечного отрезка ряда (27) для нахождения оценки плотности моменты в выражениях коэффициентов c_n заменяются их оценками.

Пример 3. Представить рядом Грама — Шарлье плотность равномерно распределенной случайной величины.

Нормируя случайную величину, получим $b = -a = \sqrt{3}$ (пример 3.3). Вычислив для этого случая моменты случайной величины и коэффициенты c_{nk}, получим разложение (25) в виде

$$f(x) = \Phi'(x) - \frac{1}{20} \Phi^{(V)}(x) + \frac{1}{105} \Phi^{(VII)}(x) + \ldots$$

На рис. 34 показано приближение к постоянной плотности одним, двумя и тремя членами разложения (кривые *1*, *2* и *3* соответственно).

Пример 4. Представить рядом Грама — Шарлье плотность дискретной случайной величины, возможными значениями которой

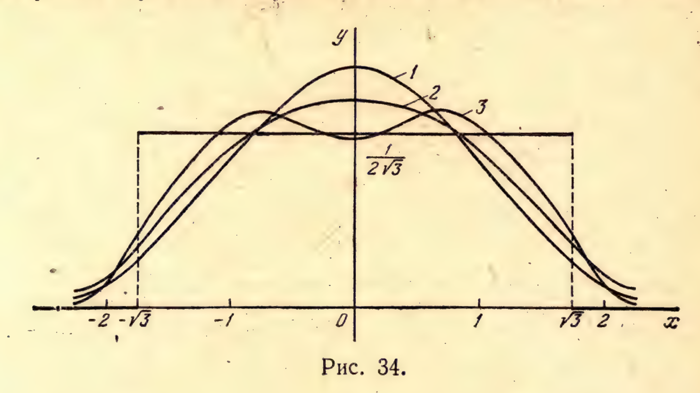


Рис. 34.

служат все целые неотрицательные числа, причем

$$P(X=m)=p_m \qquad (m=0,\ 1,\ 2,\ \ldots).$$

В данном случае естественно за «эталонное» распределение принять пуассоновское с тем же первым моментом α_1:

$$f_0(x)=\sum_{m=0}^{\infty}\frac{\alpha_1^m}{m!}e^{-\alpha_1}\,\delta(x-m).$$

Легко проверить, что ортонормальные полиномы $p_n(x)$ выражаются через моменты α_1^0, α_2^0, ... эталонного распределения $f_0(x)$ и моменты α_1, α_2, ... распределения $f(x)$ при любых $f_0(x)$ и $f(x)$ формулой

$$p_n(x)=\frac{(-1)^n\,\Delta_{2n}(x)}{\sqrt{\Delta_{2n-2}\Delta_{2n}}} \qquad (n=1,\ 2,\ \ldots),$$

где

$$\Delta_{2n}(x)=\begin{vmatrix} 1 & x & \ldots & x^n \\ 1 & \alpha_1^0 & \ldots & \alpha_n^0 \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_{n-1}^0 & \alpha_n^0 & \ldots & \alpha_{2n-1}^0 \end{vmatrix} \qquad (n=1,\ 2,\ \ldots),$$

$$\Delta_{2p}=\begin{vmatrix} 1 & \alpha_1^0 & \ldots & \alpha_p^0 \\ \alpha_1^0 & \alpha_2^0 & \ldots & \alpha_{p+1}^0 \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_p^0 & \alpha_{p+1}^0 & \ldots & \alpha_{2p}^0 \end{vmatrix} \qquad (p=1,\ 2,\ \ldots).$$

В соответствии с (17) коэффициент c_n получается заменой в определителе $\Delta_{2n}(x)$ степеней аргумента x соответствующими моментами.

В результате получаем для $f(x)$ разложение

$$f(x)=f_0(x)\left[1+\sum_{n=1}^{\infty}\frac{\Delta_{2n}(\alpha)\,\Delta_{2n}(x)}{\Delta_{2n-2}\Delta_{2n}}\right],$$

где

$$\Delta_{2n}(\alpha)=\begin{vmatrix} 1 & \alpha_1 & \ldots, & \alpha_n \\ 1 & \alpha_1^0 & \ldots & \alpha_n^0 \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_{n-1}^0 & \alpha_n^0 & \ldots & \alpha_{2n-1}^0 \end{vmatrix} \qquad (n=1, 2, \ldots).$$

В частности, для пуассоновской $f_0(x)$, приняв $\alpha_1^0=\alpha_1$ и выразив α_2^0, α_3^0, ... через α_1, получаем

$$f(x)=\sum_{m=0}^{\infty}\frac{\alpha_1^m}{m!}e^{-\alpha_1}\left[1+\frac{\alpha_2-\alpha_1-\alpha_1^2}{2}-\frac{(\alpha_2-\alpha_1-\alpha_1^2)(1+2\alpha_1)}{2\alpha_1^2}m+\right.$$
$$\left.+\frac{\alpha_2-\alpha_1-\alpha_1^2}{2\alpha_1^2}m^2+\ldots\right]\delta(x-m),$$

где

$$\alpha_n=\sum_{m=1}^{\infty}m^n p_m \qquad (n=1, 2, \ldots).$$

Пример 5. Найти приближенное выражение плотности для экспериментальных данных примера 1, ограничиваясь в формуле (27) первыми тремя слагаемыми.

В условиях примера 1 находим выборочные среднее и центральные моменты второго, третьего и четвертого порядков $\bar{x}=18{,}2$, $\mu_2^*=2{,}80\cdot 10^4$, $\mu_3^*=3{,}4\cdot 10^5$, $\mu_4^*=15{,}17\cdot 10^9$ *), после чего по формулам (3.83) и (3.84) находим оценки асимметрии и эксцесса:

$$\hat{\gamma}_1=\frac{3{,}4\cdot 10^5}{2{,}80^{3/2}\cdot 10^6}\approx 0{,}072, \quad \hat{\gamma}_2=\frac{15{,}17\cdot 10^8}{2{,}80^2\cdot 10^8}-3\approx -1{,}07.$$

Далее находим коэффициенты c_3 и c_4: $c_3=\hat{\gamma}_1/\sqrt{3!}\approx 0{,}030$; $c_4=\hat{\gamma}_2/\sqrt{4!}\approx -0{,}22$. Подставив эти значения и значение $\hat{\sigma}_x^*=\sqrt{\hat{\mu}_2^*}=167$ в (27), получаем приближенное выражение плотности:

$$\hat{f}(x)=0{,}00598\left[\Phi'\left(\frac{x-18{,}2}{167}\right)-0{,}012\Phi^{(IV)}\left(\frac{x-18{,}2}{167}\right)-\right.$$
$$\left.-0{,}044\Phi^{(V)}\left(\frac{x-18{,}2}{167}\right)\right].$$

График этой плотности показан на рис. 32 сплошной линией. Для сравнения штриховой линией показан график нормальной плотности с теми же первым и вторым моментами.

*) Для вычисления выборочных средних и центральных моментов второго, третьего и четвертого порядков по группированной выборке в [94] служит программа MOMEN, использующая для вычисления частот программу TABI.

2.4. Ряд Эджуорта. Разложение плотности по ортогональным полиномам, в частности ряд Грама—Шарлье, обладает тем свойством, что для вычисления коэффициента c_n необходимо знание моментов случайной величины до порядка n включительно. Между тем момент порядка n часто входит в выражение коэффициента c_n с таким малым коэффициентом, что практически не влияет на величину c_n. Наиболее характерен в этом отношении случай, когда величина X представляет собой сумму большого числа слагаемых. Согласно предельной теореме § 5.4 распределение величины X в этом случае при достаточно общих условиях близко к нормальному. Если представить ее плотность рядом (27), то все члены ряда, кроме первого, будут стремиться к нулю при неограниченном возрастании числа слагаемых в составе величины X. В этом случае важно получить такое разложение плотности, в котором объединены все члены одного порядка относительно n. Чтобы решить эту задачу, рассмотрим случай, когда величина X представляет собой сумму одинаково распределенных независимых случайных величин $X_1, \ldots, X_n$, имеющих моменты всех порядков.

Пусть $\mu_2 = \sigma^2$, $\mu_3, \ldots$ — моменты случайной величины $X = X_1 + \ldots + X_n$, $\mu_2' = \sigma_1^2$, $\mu_3', \ldots$ — моменты каждой из случайных величин $X_1, \ldots, X_n$. Выразим моменты μ_2, $\mu_3, \ldots$ через μ_2', $\mu_3', \ldots$ В силу независимости величин $X_1, \ldots, X_n$ получаем

$$\begin{aligned}
&\mu_2 = \sigma^2 = n\mu_2' = n\sigma_1^2, \quad \mu_3 = n\mu_3', \\
&\mu_4 = n\mu_4' + 3n(n-1)\sigma_1^4, \\
&\mu_5 = n\mu_5' + 10n(n-1)\sigma_1^2\mu_3', \\
&\mu_6 = n\mu_6' + 10n(n-1)\mu_3'^2 + 15n(n-1)\sigma_1^2\mu_4' + \\
&\qquad\qquad + 15n(n-1)(n-2)\sigma_1^6, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned} \tag{28}$$

Подставив эти выражения в (26), находим

$$\begin{aligned}
&c_3 = \frac{1}{\sqrt{3!}} \frac{\mu_3'}{\sigma_1^3 n^{1/2}}, \quad c_4 = \frac{1}{\sqrt{4!}} \left(\frac{\mu_4'}{\sigma_1^4} - 3 \right) \frac{1}{n}, \\
&c_5 = \frac{1}{\sqrt{5!}} \left(\frac{\mu_5'}{\sigma_1^5} - 10 \frac{\mu_3'}{\sigma_1^3} \right) \frac{1}{n^{3/2}}, \\
&c_6 = \frac{10}{\sqrt{6!}} \frac{\mu_3'^2}{\sigma_1^6 n} + \frac{1}{\sqrt{6!}} \left(\frac{\mu_6'}{\sigma_1^6} - 15 \frac{\mu_4'}{\sigma_1^4} - 10 \frac{\mu_3'^2}{\sigma_1^6} + 30 \right) \frac{1}{n^2}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

После подстановки этих выражений в (27) и объединения величин одного порядка малости относительно n^{-1} будем иметь

$$f(x)=\frac{1}{\sigma}\left\{\Phi'\left(\frac{x-m_x}{\sigma}\right)-\frac{\mu_3'}{3!\sigma_1^3}\Phi^{(IV)}\left(\frac{x-m_x}{\sigma}\right)n^{-1/2}+\right.$$
$$+\left[\frac{1}{4!}\left(\frac{\mu_4'}{\sigma_1^4}-3\right)\Phi^{(V)}\left(\frac{x-m_x}{\sigma}\right)+\frac{10\mu_3'^2}{6!\sigma_1^6}\Phi^{(VII)}\left(\frac{x-m_x}{\sigma}\right)\right]n^{-1}-$$
$$-\left[\frac{1}{5!}\left(\frac{\mu_5'}{\sigma_1^5}-10\frac{\mu_3'}{\sigma_1^3}\right)\Phi^{(VI)}\left(\frac{x-m_x}{\sigma}\right)+\right.$$
$$+\frac{35}{7!}\frac{\mu_3'}{\sigma_1^3}\left(\frac{\mu_4'}{\sigma_1^4}-3\right)\Phi^{(VIII)}\left(\frac{x-m_x}{\sigma}\right)+$$
$$\left.\left.+\frac{280}{9!}\frac{\mu_3'^3}{\sigma_1^9}\Phi^{(X)}\left(\frac{x-m_x}{\sigma}\right)\right]n^{-3/2}+\ldots\right\}. \quad (29)$$

Отсюда видно, что для вычисления $f(x)$ с точностью порядка $O(n^{-2})$ достаточно вычислить коэффициенты c_6, c_7, c_8 и c_9, учитывая только моменты не выше четвертого порядка. Подставив в (29) выражения моментов σ_1^2, μ_3', μ_4', ... через σ^2, μ_3, μ_4, ... из (28), приведем (29) к виду

$$f(x)=\frac{1}{\sigma}\left\{\Phi'\left(\frac{x-m_x}{\sigma}\right)-\frac{1}{3!}\frac{\mu_3}{\sigma^3}\Phi^{(IV)}\left(\frac{x-m_x}{\sigma}\right)+\right.$$
$$+\left[\frac{1}{4!}\left(\frac{\mu_4}{\sigma^4}-3\right)\Phi^{(V)}\left(\frac{x-m_x}{\sigma}\right)+\frac{10}{6!}\left(\frac{\mu_3}{\sigma^3}\right)^2\Phi^{(VII)}\left(\frac{x-m_x}{\sigma}\right)\right]-$$
$$-\left[\frac{1}{5!}\left(\frac{\mu_5}{\sigma^5}-10\frac{\mu_3}{\sigma^3}\right)\Phi^{(VI)}\left(\frac{x-m_x}{\sigma}\right)+\right.$$
$$+\frac{35}{7!}\left(\frac{\mu_4}{\sigma^4}-3\right)\Phi^{(VIII)}\left(\frac{x-m_x}{\sigma}\right)+$$
$$\left.\left.+\frac{280}{9!}\left(\frac{\mu_3}{\sigma^3}\right)^3\Phi^{(X)}\left(\frac{x-m_x}{\sigma}\right)\right]+\ldots\right\}, \quad (30)$$

где каждое из выражений в квадратных скобках содержит только величины одного порядка малости относительно n^{-1}.

Разложение (30) называется *рядом Эджуорта* [126]. Оно представляет собой асимптотическое разложение $f(x)$*) по степеням $n^{-1/2}$, причем, как показал Крамер,

---

*) Ряд $\sum a_k\varphi_k(z)$ называется *асимптотическим разложением* функции $f(z)$ при $z\longrightarrow z_0$, если при любом n $\varphi_n(z)\longrightarrow 0$ при $z\longrightarrow z_0$ и

$$f(x)-\sum_{k=1}^{n}a_k\varphi_k(z)=o(\varphi_n(z)).$$

остаточный член имеет порядок первого отброшенного члена [45].

Если ограничиться в (30) членами порядка n^{-1} и учесть, что величины μ_3/σ^3 и $(\mu_4/\sigma^4)-3$ согласно (3.83) и (3.84) представляют собой соответственно асимметрию γ_1 и эксцесс γ_2 распределения случайной величины X, то получим следующее приближенное выражение для плотности величины X:

$$f(x)=\frac{1}{\sigma}\left\{\Phi'\left(\frac{x-m_x}{\sigma}\right)-\frac{\gamma_1}{3!}\Phi^{(IV)}\left(\frac{x-m_x}{\sigma}\right)+\right.$$
$$\left.+\frac{\gamma_2}{4!}\Phi^{(V)}\left(\frac{x-m_x}{\sigma}\right)+\frac{10\gamma_1^2}{6!}\Phi^{(VII)}\left(\frac{x-m_x}{\sigma}\right)\right\}. \qquad (31)$$

Этой формулой часто пользуются на практике.

Аналогично можно получить ряд Эджуорта для плотности векторной случайной величины [20].

2.5. Представление плотности линейной комбинацией данных функций. При пользовании конечными отрезками разложений по ортогональным полиномам, как мы видели в примере 3, иногда получаются малые отрицательные значения плотности. Это естественное следствие того, что любой конечный отрезок ряда дает лишь приближенное значение функции, представляемой данным рядом.

Перейдем к общему способу приближенного представления плотности линейной комбинацией заданных функций. Пусть $f(x)$ — плотность случайной величины X, $\varphi_1(x), \ldots, \varphi_N(x)$ — заданные функции. Задача приближения плотности $f(x)$ линейной комбинацией функций $\varphi_1(x), \ldots, \varphi_N(x)$ сводится к определению коэффициентов $c_1, \ldots, c_N$ в приближенной формуле

$$f(x)\approx c_1\varphi_1(x)+\ldots+c_N\varphi_N(x). \qquad (32)$$

Эта задача может быть решена различными способами. Она решается стандартными методами теории приближения функций, если $f(x)$ известна. Однако в теории веро-

Очевидно, что всякий степенной ряд, сходящийся к функции $f(x)$, является ее асимптотическим разложением при $z\to 0$. Однако асимптотическое разложение может и не сходиться ни при каком z. Асимптотические разложения, наряду со сходящимися рядами, широко применяются для вычисления значений функций и для решения практических задач. При этом асимптотические разложения по положительным степеням аргумента дают достаточную точность только при достаточно малых значениях аргумента, а асимптотические разложения по отрицательным степеням аргумента — только при достаточно больших значениях аргумента.

ятностей обычно приходится решать эту задачу, когда плотность $f(x)$ не известна. В этом случае приходится пользоваться специальными приемами.

Если известно $N-1$ моментов величины X, то, приравняв эти моменты соответствующим моментам, вычисленным с помощью выражения (32), получим $N-1$ уравнений, связывающих коэффициенты $c_1, \ldots, c_N$. Добавив к этим уравнениям очевидное соотношение

$$c_1 \int \varphi_1(x)\,dx + \ldots + c_N \int \varphi_N(x)\,dx = 1,$$

получим N линейных алгебраических уравнений для определения $c_1, \ldots, c_N$.

В частном случае скалярной величины X, зная ее моменты $\alpha_1, \ldots, \alpha_{N-1}$, получаем уравнения

$$\sum_{k=1}^{N} c_k \int x^r \varphi_k(x)\,dx = \alpha_r \quad (r = 0, 1, \ldots, N-1;\ \alpha_0 = 1).$$

В случае векторной величины X следует выбрать N равным числу всех известных моментов, увеличенному на 1.

В случае, когда требуется найти оценку плотности $f(x)$ вида (32) по результатам опытов, задача сводится к оцениванию неизвестных коэффициентов $c_1, \ldots, c_N$.

Эта задача решается любым методом нахождения оценок неизвестных параметров распределений. В частности, оценки коэффициентов $c_1, \ldots, c_N$ можно найти методом максимума правдоподобия или методом моментов. В последнем случае она решается так же, как в случае известных моментов, только вместо самих моментов в уравнения для $c_1, \ldots, c_N$ войдут их оценки.

Если взять в качестве функций $\varphi_1, \ldots, \varphi_N$ некоторые плотности, то часто удается избежать отрицательных плотностей при вычислении их по формуле (32).

Можно также искать приближенное представление плотности линейной комбинацией вида (32) заданных функций $\varphi_1, \ldots, \varphi_N$, зависящих от неизвестных параметров. В этом случае задача сводится к совместному определению или оцениванию коэффициентов $c_1, \ldots, c_N$ и неизвестных параметров, входящих в выражения функций $\varphi_1, \ldots, \varphi_N$. При этом можно взять меньше слагаемых в (32), но зато уравнения для неизвестных параметров или их оценок, получаемые методом моментов, уже не будут линейными.

Пример 6. Представить распределение дискретной случайной величины примера 4 линейной комбинацией двух распределений Пуассона:

$$f(x)=\sum_{m=0}^{\infty}\left[c\,\frac{\mu^m}{m!}\,e^{-\mu}+(1-c)\,\frac{h^m\mu^m}{m!}\,e^{-h\mu}\right]\delta(x-m),$$

считая параметр μ неизвестным, а параметр $h>1$ известным.

В данном случае $c_1+c_2=1$. Поэтому мы положили $c_1=c, c_2=1-c$. Для определения c и μ применим метод моментов. Имея в виду, что первые два момента распределения Пуассона $\mu^m e^{-\mu}/m!$ равны μ и $\mu+\mu^2$, получаем уравнения

$$c\mu+(1-c)\,\mu h=\alpha_1,$$
$$c(\mu+\mu^2)+(1-c)(h\mu+h^2\mu^2)=\alpha_2.$$

Решив эти уравнения и приняв во внимание, что при $\alpha_2=\alpha_1+\alpha_1^2$ эти уравнения имеют очевидное решение $c=1$, $\mu=\alpha_1$, получим

$$c=\frac{2h-\beta(1+h)+\sqrt{\beta^2(1+h)^2-4\beta h}}{2(h-1)},$$

$$\mu=\frac{2\alpha_1}{\beta(1+h)-\sqrt{\beta^2(1+h)^2-4\beta h}},\quad \beta=\frac{\alpha_1^2}{\alpha_2-\alpha_1}.$$

§ 3. Проверка гипотез о распределениях

3.1. Задачи проверки гипотез. Из сказанного в § 6.1 ясно, что распределение случайной величины принципиально невозможно точно определить по результатам опытов. Полученные экспериментально оценки распределения дают возможность только строить различные гипотезы о распределении случайной величины, например гипотезу о том, что она распределена нормально. Поэтому возникает задача проверки гипотез. Эта задача состоит в том, чтобы определить, насколько хорошо согласуется та или иная гипотеза о распределении случайной величины с полученными экспериментальными данными. Ясно, что эта задача тесно связана с задачей определения доверительных областей для плотности или функции распределения. Однако она имеет и некоторые особенности. Дело в том, что по той же выборке, по которой проверяется гипотеза о распределении, обычно оцениваются и некоторые параметры этого распределения. Так, например, проверяя гипотезу о нормальном распределении, по той же выборке обычно оценивают математическое ожидание и ковариационную матрицу (дисперсию в случае одномерного распределения) случайной величины. Вследствие

этого гипотетическое распределение оказывается само случайным — функцией случайных результатов опытов. Это и отличает задачу проверки гипотез о распределении от задачи определения доверительных областей для распределений. И только в отдельных частных случаях может возникнуть задача проверки гипотезы о том, что случайная величина подчинена вполне определенному закону распределения, не зависящему от неизвестных параметров.

Для проверки гипотез о распределении применяются различные критерии согласия. Наиболее удобным и универсальным критерием согласия является критерий χ^2 К. Пирсона. Он совершенно не зависит ни от распределения случайной величины, ни от ее размерности.

3.2. Критерий χ^2. Критерий χ^2 К. Пирсона основан на использовании в качестве меры отклонения экспериментальных данных от гипотетического распределения той же величины, которая служит для построения доверительной области для неизвестной плотности, с заменой неизвестных истинных значений вероятностей попадания в интервалы вероятностями, вычисленными по гипотетическому распределению. Предположим, что область возможных значений случайной величины разбита на r интервалов (многомерных, т. е. прямоугольников, в случае векторной величины). Пусть, как и в § 1, $\hat{P}_1, \ldots, \hat{P}_r$ — случайные частоты попадания в эти интервалы, получаемые в результате n опытов, $P_1, \ldots, P_r$ — вероятности попадания в те же интервалы, вычисленные по гипотетическому распределению. В общем случае эти вероятности являются функциями оценок неизвестных параметров, получаемых по тем же экспериментальным данным, и потому тоже являются случайными величинами. Предположим, что оценки неизвестных параметров гипотетического распределения вычисляются по той же группированной выборке, что и частоты $\hat{P}_1, \ldots, \hat{P}_r$. Тогда вероятности $P_1, \ldots, P_r$ будут некоторыми функциями частот $\hat{P}_1, \ldots, \hat{P}_r$, и вместо величины H, определяемой формулой (1), для оценки отклонения экспериментальных данных от гипотетического распределения придется взять величину

$$Z = n \sum_{\nu=1}^{r} \frac{(\hat{P}_\nu - P_\nu)^2}{P_\nu}, \tag{33}$$

где $P_1, \ldots, P_r$ — определенные функции частот $\hat{P}_1, \ldots, \hat{P}_r$.

Нейман и Е. Пирсон показали, что *если для вычисления вероятностей* $P_1, \ldots, P_r$ *применяется асимптотически эффективная и асимптотически нормальная оценка неизвестного s-мерного параметра гипотетического распределения по группированной выборке, то величина* Z, *определяемая формулой* (33), *в пределе при* $n \to \infty$ *имеет* χ^2*-распределение с* $r-s-1$ *степенями свободы* [65, 45].

Пользуясь этой теоремой, можно оценивать расхождение экспериментальных данных с гипотетическим распределением с помощью таблиц χ^2-распределения (таблица 5). Выберем достаточно малую вероятность p, чтобы событие с такой вероятностью можно было считать практически невозможным, и определим χ_p^2 из уравнения

$$P(Z \geqslant \chi_p^2) = \int_{\chi_p^2}^{\infty} p_k(u)\,du = p, \qquad k = r-s-1.$$

Если реализация $\zeta = \chi^2$ величины Z, полученная в результате опытов, превосходит χ_p^2 или равна χ_p^2, $\zeta = \chi^2 \geqslant \chi_p^2$, то гипотетическое распределение считают не согласующимся с экспериментальными данными, так как при этом распределении практически невозможно получить при одной выборке $\zeta = \chi^2 \geqslant \chi_p^2$. Вероятность такого события при большом числе опытов n приближенно равна p, т. е. пренебрежимо мала. В этом случае говорят, что имеет место *значимое отклонение* экспериментальных данных от гипотетического распределения. Если же $\zeta = \chi^2 < \chi_p^2$, то считают, что гипотетическое распределение не противоречит экспериментальным данным, согласуется с ними.

Величина χ_p^2 называется *100p-процентным уровнем значимости* отклонения выборки от гипотетического распределения. В приложениях обычно пользуются 5-, 1- и 0,1-процентными уровнями значимости, в зависимости от характера задачи.

Для дополнительной проверки согласованности экспериментальных данных с гипотетическим распределением полезно вычислить вероятность того, что при данном гипотетическом распределении величина Z окажется больше полученной в результате опытов ее реализации $\zeta = \chi^2$, $P(Z > \chi^2)$. Эту вероятность можно также приближенно определить по таблице 5. Чем больше эта вероятность, тем лучше согласуется выборка с гипотетическим распределением, тем меньше значимость полученного расхождения выборки с гипотетическим распределением. Дейст-

вительно, если вероятность $P(Z > \chi^2)$ велика, то при повторении данной серии опытов в случае справедливости выбранной гипотезы о распределении часто будут получаться значения величины Z еще большие, чем полученное в результате опытов значение $\zeta = \chi^2$.

Обратим внимание на то, что, получив $\zeta = \chi^2 < \chi_p^2$ и даже получив высокую вероятность $P(Z > \chi^2)$, мы не делаем определенного вывода, что выбранная гипотеза о распределении справедлива, а говорим лишь, что эта гипотеза не противоречит полученным результатам опытов, что она согласуется с ними, вследствие чего ее можно принять. Чтобы получить достаточно веское доказательство того, что случайная величина действительно подчинена гипотетическому закону распределения, необходимо повторить данную серию опытов достаточно большое число раз и убедиться в том, что полученное согласование гипотезы с результатами опытов устойчиво.

Если $k = r - s - 1 > 30$, то величину $X = \sqrt{Z}$ можно считать распределенной нормально $N(\sqrt{k - 1/2},\ 1/2)$ и для вычисления величины χ_p^2 воспользоваться формулой

$$P(X \geqslant \chi_p) = \frac{1}{2} - \Phi(\chi_p \sqrt{2} - \sqrt{2k - 1}) = p$$

и таблицей функции Лапласа (табл. 1).

Подчеркнем, что при вычислении полученного в результате опытов значения $\zeta = \chi^2$ необходимо учитывать также бесконечные интервалы с нулевыми частотами и отличными от нуля при гипотетическом распределении вероятностями.

Пример 7. В условиях примера 1 проверить гипотезу о нормальном распределении случайной величины.

В данном случае с учетом бесконечных 11-го $(-\infty, 385)$ и 12-го $(415, \infty)$ интервалов имеем $r = 12$. Так как по той же группированной выборке мы оцениваем два параметра m_x и D_x, то $s = 2$ и $k = r - s - 1 = 9$. По таблице 5 находим для $k = 9$ и $p = 0{,}05$ 5%-ный уровень значимости $\chi_p^2 = 16{,}92$. Для вычисления полученного значения χ^2 находим по гипотетическому нормальному распределению при $m_x = \bar{x} = 18{,}2$ вероятности попадания в интервалы Δx_ν. Полученные значения частот и вероятностей приведены в таблице.

По этим данным находим по формуле (33) $\chi^2 = 4{,}157$. Так как это значение χ^2 значительно меньше критического значения χ_p^2, то можно считать, что гипотеза о нормальном распределении случайной величины не противоречит результатам опытов.

Вычислим еще вероятность значения меры расхождения Z большего, чем значение $\chi^2 = 4{,}157$, полученное в результате опытов. По

ν	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{p}_\nu \cdot 10^4$	200	600	800	1600	1900	1700	1400	800	700	300	0	0
$p_\nu \cdot 10^4$	190	467	919	1440	1826	1837	1487	959	496	206	82	91

таблице 5 при $k=9$ и $\chi_p^2=4,157$ находим $P(Z > 4,157)=0,908$. Таким образом, при справедливости нашей гипотезы о распределении можно с вероятностью $\approx 0,91$ получить еще большее расхождение χ^2, чем то, которое получилось в результате опытов. Это служит дополнительным основанием для принятия гипотезы о том, что случайная величина X распределена $N(18,2,\ 168)$.

Пример 8. Для тех же экспериментальных данных проверить гипотезу о том, что распределение определяется плотностью, найденной в примере 5.

Пользуясь формулой примера 5 для плотности и таблицами 1 и 2 функции Лапласа и ее производных, находим вероятности попадания в интервалы. Результаты приведены в таблице.

ν	1	2	3	4	5	6	7	8	9	10	11	12
$p_\nu \cdot 10^4$	209	582	1073	1494	1624	1621	1440	1062	614	243	0,3	35

Подставив эти вероятности и соответствующие частоты из примера 5 в формулу (33), находим $\chi^2=2,528$. Так как в данном случае по той же группированной выборке оцениваются четыре параметра m_x, σ_x, γ_1 и γ_2, то $s=4$ и χ^2-распределение имеет $k=r-s-1=7$ степеней свободы. По таблице 5 находим для $k=7$ и $p=0,05$ критическое значение $\chi_p^2=14,07$ и вероятность получить значение χ^2 большее, чем 2,528, $P(Z > 2,528)=0,923$. Таким образом, распределение, подобранное в примере 5, согласуется с экспериментальными данными несколько лучше, чем нормальное распределение.

Пример 9. В условиях примера 7 проверить гипотезу о распределении $N(15,\ 160^2)$.

В данном случае мы не оцениваем параметры нормального распределения по выборке, а просто задаем их. Основанием для их выбора служит только то, что точка $x=15$ является средней точкой интервала, занятого экспериментальными точками, причем 51 экспериментальное значение случайной величины меньше 15, а 49 — больше 15, и что 66% всех точек лежат в интервале (15—160, 15+160), в то время как при нормальном распределении вероятность попадания в интервал $(m_x-\sigma_x,\ m_x+\sigma_x)$ равна 0,68. Поэтому в качестве меры расхождения между гипотетическим распределением и экспериментальными данными можно принять величину H § 1. Предельным распределением этой величины служит χ^2-распределение с $r-1=11$ степенями свободы. При этом получается $\chi^2=5,221$ при 5%-ном уровне

значимости $\chi_p^2 = 19{,}68$ и $P(\mathrm{H} > \chi^2) = 0{,}9178$. Таким образом, несмотря на несколько большее расхождение между вероятностями p_ν и частотами $\hat{p}_\nu$ по сравнению с примером 7, χ^2-критерий дает не меньше оснований для принятия гипотезы о распределении $N(15,\ 160^2)$, чем для принятия гипотезы о распределении $N(18{,}2,\ 168^2)$.

3.3. Вывод предельного χ^2-распределения. Докажем теперь теорему Неймана и Е. Пирсона об асимптотическом распределении величины

$$Z = n \sum_{\nu=1}^{r} \frac{[\hat{P}_\nu - p_\nu(\hat{\Theta})]^2}{p_\nu(\hat{\Theta})} = W^T W.$$

где $\hat{\Theta} = \varphi(\hat{P})$ — оценка параметра θ, представляющая собой некоторую функцию вектора частот $\hat{P} = [\hat{P}_1 \ \ldots \ \hat{P}_r]^T$ (т. е. найденная по группированной выборке), а W — вектор с координатами $W_\nu = [\hat{P}_\nu - p_\nu(\hat{\Theta})]\sqrt{n/p_\nu(\hat{\Theta})}$ $(\nu = 1, \ \ldots, \ r)$.

▶ Для нахождения асимптотических распределений величин W и Z выразим вектор W приближенно через вектор $V = [V_1 \ \ldots \ V_r]^T$:

$$V_\nu = [\hat{P}_\nu - p_\nu(\theta)]\sqrt{n/p_\nu(\theta)} \qquad (\nu = 1, \ \ldots, \ n).$$

Предположим сначала, что $\hat{\Theta}$ представляет собой оценку максимального правдоподобия, т. е. корень уравнения (7.31), которое мы запишем в виде $\sum_{\nu=1}^{r} \hat{P}_\nu \frac{\partial \ln p_\nu(\hat{\Theta})}{\partial \hat{\Theta}} = 0.$

Поскольку $\sum p_\nu(\hat{\Theta}) = 1$, то $\sum_{\nu=1}^{r} p_\nu(\hat{\Theta}) \frac{\partial \ln p_\nu(\hat{\Theta})}{\partial \hat{\Theta}} = 0$. Вычитая это уравнение из предыдущего, получим

$$\sum_{\nu=1}^{r} [\hat{P}_\nu - p_\nu(\hat{\Theta})] \frac{\partial \ln p_\nu(\hat{\Theta})}{\partial \hat{\Theta}} = 0. \tag{34}$$

Имея в виду, что $\hat{\Theta} \to \theta$ и $\hat{P} \to p(\Theta)$ по вероятности при $n \to \infty$, будем искать решение уравнения (34) в виде разложения по степеням разности $\hat{P} - p$ или, что одно и то же, по степеням вектора V. Тогда с точностью до малых первого порядка будем иметь

$$\sum_{\nu=1}^{r} \left[\hat{P}_\nu - p_\nu - \frac{\partial^T p_\nu}{\partial \theta} (\hat{\Theta} - \theta) \right] \frac{1}{p_\nu} \frac{\partial p_\nu}{\partial \theta} = 0,$$

или, после умножения на $\sqrt{n}$,

$$\sum_{\nu=1}^{r} \frac{1}{\sqrt{p_\nu}} \frac{\partial p_\nu}{\partial \theta} V_\nu - \sqrt{n} \left(\sum_{\nu=1}^{r} \frac{1}{p_\nu} \frac{\partial p_\nu}{\partial \theta} \frac{\partial^T p_\nu}{\partial \theta} \right) (\hat{\Theta} - \theta) = 0.$$

Вводя матрицу

$$C = \left[\frac{1}{\sqrt{p_1}} \frac{\partial p_1}{\partial \theta} \cdots \frac{1}{\sqrt{p_r}} \frac{\partial p_r}{\partial \theta} \right],$$

можем переписать полученное уравнение в виде

$$CV - \sqrt{n} CC^T (\hat{\Theta} - \theta) = 0.$$

Если матрица $D = [\partial p_1/\partial\theta \ \ldots \ \partial p_r/\partial\theta]$ имеет ранг s, то и C имеет ранг s. Поэтому матрица CC^T тоже имеет ранг s и, следовательно, существует обратная матрица $(CC^T)^{-1}$ и полученное уравнение имеет решение

$$\hat{\Theta} = \theta + \frac{1}{\sqrt{n}} (CC^T)^{-1} CV. \tag{35}$$

Эта формула с точностью до величин порядка $o(n^{-1/2})$ определяет и решение исходного уравнения (34).

Пользуясь формулой (35), можно выразить вектор W через V с точностью до величин порядка $o(n^{-1/2})$. Для этого заметим, что с точностью до величин порядка $\sqrt{n}\, o(|\hat{\Theta} - \theta|)$

$$W_\nu = \sqrt{n} \frac{\hat{P}_\nu - p_\nu(\hat{\Theta})}{\sqrt{p_\nu(\hat{\Theta})}} = \sqrt{\frac{n}{p_\nu}} \left[\hat{P}_\nu - p_\nu - \frac{\partial^T p_\nu}{\partial \theta} (\hat{\Theta} - \theta) \right].$$

Подставив сюда выражение (35), получим с точностью до величин порядка $\sqrt{n}\, o(n^{-1/2}) = O(n^{-1/2})$

$$W_\nu = V_\nu - \frac{1}{\sqrt{p_\nu}} \frac{\partial^T p_\nu}{\partial \theta} (CC^T)^{-1} CV,$$

или, в матричной форме,

$$W = [I - C^T (CC^T)^{-1} C] V. \tag{36}$$

Итак, вектор W выражается через V асимптотической формулой (36). Следовательно, ковариационная матрица вектора W выражается через ковариационную матрицу вектора V, $K_v = I - qq^T$, формулой

$$K_w = [I - C^T (CC^T)^{-1} C] (I - qq^T) [I - C^T (CC^T)^{-1} C] = \\ = I - qq^T - C^T (CC^T)^{-1} C.$$

При выводе этой формулы мы воспользовались соотношением $Cq=0$.

Умножив (36) слева на матрицу C, получаем $CW=0$. Это равенство дает s скалярных соотношений между координатами вектора W. Кроме того, координаты вектора W связаны скалярным соотношением $q(\hat{\Theta})^T W=0$, где $q(\hat{\Theta})$ — вектор с координатами $\sqrt{p_1(\hat{\Theta})}, \ldots, \sqrt{p_r(\hat{\Theta})}$. В пределе при $n \to \infty$ это соотношение переходит в то же соотношение, которым связаны координаты вектора V, $q^T W=0$. Таким образом, координаты вектора W связаны $s+1$ скалярными соотношениями. Это значит, что асимптотическое распределение вектора W является вырожденным нормальным распределением, сосредоточенным на $(r-s-1)$-мерном подпространстве, и, следовательно, его ковариационная матрица K_w имеет ранг $r-s-1$. Воспользуемся теперь последним результатом примера 5.33. Матрица A в этом случае представляет собой единичную матрицу I, а $B=K_w^{1/2}AK_w^{1/2}=K_w$. Таким образом, матрица B имеет ранг $r-s-1$ и удовлетворяет соотношению $K_w^2=K_w$ в силу того же равенства $Cq=0$, т. е. является матрицей проектирования на некоторое $(r-s-1)$-мерное подпространство r-мерного пространства.

В соответствии с последним результатом примера 5.33 отсюда и следует, что асимптотическим распределением величины $Z=W^T W$ служит χ^2-распределение с $r-s-1$ степенями свободы. ◀

Этот результат мы получили, предполагая, что оценка $\hat{\Theta}$ параметра θ является оценкой максимума правдоподобия. Однако этот результат верен и для случая произвольной асимптотически нормальной и асимптотически эффективной оценки $\hat{\Theta}$, так как все такие оценки с точностью до величины порядка $o(n^{-1/2})$ совпадают и определяются формулой (35) (в противном случае отношения дисперсий этих оценок не могли бы сходиться к 1 при $n \to \infty$).

3.4. Оценивание параметров распределений методом минимума χ^2. Изложенный метод проверки гипотез о распределениях естественно наводит на мысль искать такие оценки неизвестных параметров гипотетического распределения, при которых полученная реализация $\zeta=\chi^2$ меры расхождения Z экспериментальных данных с гипотетическим распределением имеет минимальное значение. Так

мы приходим к новому методу нахождения оценок параметров распределения — *методу минимума* χ^2.

В случае, когда вероятности $p_1, \ldots, p_r$ попадания в выбранные интервалы имеют непрерывные производные по всем координатам вектора θ, метод минимума χ^2 дает для оценки параметра θ уравнение

$$-2n\sum_{\nu=1}^{r}\frac{\hat{p}_\nu - p_\nu}{p_\nu}\frac{\partial p_\nu}{\partial\theta} - n\sum_{\nu=1}^{r}\frac{(\hat{p}_\nu - p_\nu)^2}{p_\nu^2}\frac{\partial p_\nu}{\partial\theta} = 0,$$

где, как всегда, $\partial/\partial\theta$ представляет собой вектор с координатами $\partial/\partial\theta_1, \ldots, \partial/\partial\theta_s$. Учитывая, что $n\hat{p}_\nu = m_\nu$ представляет собой число экспериментальных точек в ν-м интервале ($\nu = 1, \ldots, r$) и что $\sum \partial p_\nu/\partial\theta = 0$ вследствие равенства $\sum p_\nu = 1$, перепишем полученное уравнение в виде

$$\sum_{\nu=1}^{r} m_\nu \frac{\partial \ln p_\nu}{\partial\theta} + \frac{1}{2}\sum_{\nu=1}^{r}\frac{(m_\nu - np_\nu)^2}{np_\nu}\frac{\partial \ln p_\nu}{\partial\theta} = 0.$$

Вторая сумма здесь обычно мала по сравнению с первой. Если пренебречь второй суммой, то уравнение минимума χ^2 совпадет с уравнением максимума правдоподобия (7.31). Поэтому оценки минимума χ^2 обычно близки к оценкам максимума правдоподобия.

Метод минимума χ^2, как и метод максимума правдоподобия, дает при весьма общих условиях асимптотически эффективную и асимптотически нормальную оценку параметра θ (как в случае скалярного, как и в случае векторного θ).

3.5. Другие методы проверки гипотез о распределениях. Кроме метода χ^2, можно применять другие методы проверки гипотез о распределениях. Крамером [43], Мизесом [60] и Н. В. Смирновым [97] предложены критерии согласия, основанные на мерах расхождения, зависящих от разности $\hat{F}(x) - F(x)$ — отклонения оценки $\hat{F}(x)$ функции распределения от гипотетической функции распределения $F(x)$. К критериям такого рода относится критерий Колмогорова, основанный на предельном распределении величины $U = \sqrt{n}\max|\hat{F}(x) - F(x)|$ при $n \to \infty$ в случае полностью заданной гипотетической функции распределения $F(x)$ скалярной случайной величины X. Пользуясь таблицей предельной функции распределения Колмогорова $K(u)$ (таблица 7), можно про-

верять гипотезы об одномерных распределениях совершенно так же, как при применении χ^2-критерия. Задав достаточно малую вероятность p, можно найти по таблице функции $K(u)$ $100p\%$-ный уровень значимости u_p, определяемый уравнением $P(U \geqslant u_p) = 1 - K(u_p) = p$. Если полученное по результатам опытов значение u меньше критического значения u_p, то можно считать, что результаты опытов не противоречат гипотезе, что функция распределения наблюдаемой величины X есть $F(x)$. Если же $u \geqslant u_p$, то эту гипотезу следует отвергнуть*).

Аналогично можно пользоваться критерием Н. В. Смирнова, основанным на найденном им предельном распределении величины $nM[|\hat{F}(X) - F(X)|^2 | x_1, \ldots, x_n]$, которое в случае скалярной величины X не зависит от $F(X)$. Однако теория проверки гипотез, основанная на таких критериях, пока еще очень слабо разработана. Поэтому χ^2-критерий остается пока что единственным практически приемлемым критерием для проверки гипотез о распределениях, заданных лишь с точностью до неизвестных, подлежащих оцениванию по той же выборке параметров, и гипотез о многомерных распределениях.

3.6. Проверка гипотез о независимости случайных величин. Вследствие универсальности χ^2-критерия и его применимости к многомерным распределениям он может служить и для проверки множества гипотез других типов.

Покажем, как можно с помощью χ^2-критерия проверять гипотезу о независимости случайных величин. Предположим, что область значений величины X разбита на r_1 интервалов, а область значений величины Y — на r_2 интервалов. Пусть $\hat{P}_{ij}$ и p_{ij} — случайная частота и вероятность попадания вектора $[X^T \ Y^T]^T$ в пересечение i-го интервала значений X и j-го интервала значений Y ($i = 1, \ldots, r_1$; $j = 1, \ldots, r_2$). Если X и Y независимы, то $p_{ij} = p_{i\cdot} p_{\cdot j}$, где $p_{i\cdot}$ и $p_{\cdot j}$ — вероятности попадания X в i-й интервал и Y в j-й ($i = 1, \ldots, r_1$; $j = 1, \ldots, r_2$). Вероятности $p_{i\cdot}$ и $p_{\cdot j}$ можно рассматривать как $r_1 + r_2 - 2$ неизвестных параметров распределения вектора $[X^T \ Y^T]^T$ (с учетом соотношений $\sum p_{i\cdot} = \sum p_{\cdot j} = 1$). Согласно результатам примеров 7.3 и 5.48 эффективными асимпто-

---

*) Для проверки гипотез о распределениях по критерию Колмогорова в [94] служит программа KOLMO, использующая для вычисления функции распределения $K(u)$ программу SMIRN.

тически нормальными оценками вероятностей $p_{i\cdot}$ и $p_{\cdot j}$ могут служить соответствующие частоты

$$\hat{P}_{i\cdot} = \sum_{j=1}^{r_2} \hat{P}_{ij}, \quad \hat{P}_{\cdot j} = \sum_{i=1}^{r_1} \hat{P}_{ij}.$$

Поэтому величина

$$Z = n \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \frac{(\hat{P}_{ij} - \hat{P}_{i\cdot}\hat{P}_{\cdot j})^2}{\hat{P}_{i\cdot}\hat{P}_{\cdot j}} \tag{37}$$

имеет асимптотическое χ^2-распределение с $k = r_1 r_2 - (r_1 + r_2 - 2) - 1 = (r_1 - 1)(r_2 - 1)$ степенями свободы. Это дает возможность проверять гипотезу о независимости двух величин, как скалярных, так и векторных *).

Совершенно так же проверяются гипотезы о независимости любого числа случайных величин.

3.7. Проверка гипотез о совпадении распределений. К другому типу гипотез, проверяемых с помощью χ^2-критерия, относятся гипотезы о совпадении распределений в нескольких сериях опытов, так называемые гипотезы об однородности статистических данных.

Предположим, что в результате N независимых последовательностей опытов, содержащих $n_1, \ldots, n_N$ наблюдений, $n_1 + \ldots + n_N = n$, получены частоты попадания наблюдаемой величины X в интервалы Δ_ν $(\nu = 1, \ldots, r)$, на которые разбита область ее возможных значений. Требуется проверить гипотезу о совпадении распределений величины X (или N различных наблюдаемых величин) в этих N последовательностях опытов.

Пусть $\hat{P}_{\mu\nu}$ — случайная частота попадания наблюдаемой величины в ν-й интервал в μ-й последовательности опытов, p_ν — вероятность попадания в ν-й интервал $(\nu = 1, \ldots, r;\ \mu = 1, \ldots, N)$. Так как сумма независимых величин с χ^2-распределениями имеет χ^2-распределение с суммарным числом степеней свободы (пример 5.31), то случайная величина

$$H = \sum_{\mu=1}^{N} n_\mu \sum_{\nu=1}^{r} \frac{(\hat{P}_{\mu\nu} - p_\nu)^2}{p_\nu} \tag{38}$$

при данных вероятностях $p_1, \ldots, p_r$ имеет асимптоти-

---

*) В [94] для проверки гипотезы о независимости двух случайных величин по χ^2-критерию служит программа CHISQ.

ческое χ^2-распределение с $N(r-1)$ степенями свободы, если распределение наблюдаемой величины одно и то же во всех сериях опытов. Это дает возможность проверять гипотезу о том, что во всех N последовательностях опытов наблюдаемая величина имеет одно и то же распределение, для которого вероятности попадания в интервалы имеют данные значения $p_1, \ldots, p_r$.

Если гипотетическое распределение зависит от неизвестного s-мерного параметра θ, который оценивается по тем же результатам опытов, то можно пользоваться следующей теоремой, которая доказывается путем очевидного обобщения выкладок п. 3.3: *если пользоваться асимптотически эффективными и асимптотически нормальными оценками параметра θ по объединенной группированной выборке, то случайная величина*

$$Z = \sum_{\mu=1}^{N} n_\mu \sum_{\nu=1}^{r} \frac{(\hat{P}_{\mu\nu} - P_\nu)^2}{P_\nu} \tag{39}$$

имеет асимптотическое χ^2-распределение с $N(r-1)-s$ степенями свободы. Эта теорема дает возможность проверять гипотезы о совпадении распределений в нескольких сериях опытов с одновременной оценкой их параметров.

За неизвестные параметры распределения в таких задачах можно принять, в частности, вероятности $p_1, \ldots, p_{r-1}$, $p_r = 1 - p_1 - \ldots - p_{r-1}$. В этом случае $s = r - 1$, а за оценки P_ν вероятностей p_ν следует принять соответствующие частоты, вычисленные по объединенной выборке:

$$P_\nu = \frac{1}{n} \sum_{\mu=1}^{N} n_\mu \hat{P}_{\mu\nu} \quad (\nu = 1, \ldots, r), \qquad \sum_{\nu=1}^{r} P_\nu = 1. \tag{40}$$

Величина Z, определяемая формулой (39), имеет в этом случае $k = N(r-1) - (r-1) = (N-1)(r-1)$ степеней свободы. Этим критерием можно, в частности, пользоваться для проверки гипотез о совпадении дискретных распределений.

В частном случае при $r = 2$ отсюда вытекает критерий для проверки гипотез о совпадении вероятностей события в нескольких сериях опытов. Полагая в (39) и (40) $r = 2$, $\hat{P}_{\mu1} = \hat{P}_\mu$, $\hat{P}_{\mu2} = 1 - \hat{P}_\mu$,

$$P_1 = \hat{P} = \frac{1}{n} \sum_{\mu=1}^{N} n_\mu \hat{P}_\mu, \qquad P_2 = 1 - \hat{P} = \hat{Q},$$

получим

$$Z = \sum_{\mu=1}^{N} n_\mu \frac{(\hat{P}_\mu - \hat{P})^2}{\hat{P}\hat{Q}}. \qquad (41)$$

Эта величина имеет асимптотическое χ^2-распределение с $N-1$ степенями свободы. С помощью этой величины Z можно проверять гипотезы о совпадении вероятностей интересующего нас события в нескольких сериях опытов.

В случае скалярной наблюдаемой величины X с непрерывной функцией распределения для проверки гипотезы о совпадении распределений в двух выборках можно пользоваться критерием Н. В. Смирнова, аналогичным критерию Колмогорова [98]. Пусть $\hat{F}_1(x)$ и $\hat{F}_2(x)$ — выборочные функции распределения наблюдаемой величины в двух сериях опытов, состоящих из n_1 и n_2 опытов соответственно. Н. В. Смирнов показал, что если функция распределения наблюдаемой величины X непрерывна, то величина

$$U = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \max |\hat{F}_1(x) - \hat{F}_2(x)| \qquad (42)$$

имеет то же предельное распределение Колмогорова при $n_1, n_2 \to \infty$, что и каждая из величин $\sqrt{n_1} \max |\hat{F}_1(x) - F(x)|$ и $\sqrt{n_2} \max |\hat{F}_2(x) - F(x)|$. Это дает возможность проверять гипотезу о совпадении функций распределения в двух сериях опытов с помощью предельной функции распределения Колмогорова $K(u)$ *).

§ 4. Метод статистического моделирования

4.1. Задачи статистического моделирования. Непосредственное экспериментальное изучение сложных случайных явлений часто требует чрезмерно больших затрат средств и времени, а иногда и принципиально невозможно. Так, например, экспериментальное изучение функционирования сложной системы невозможно до тех пор, пока система не создана и не изготовлена. Между тем необходимо еще на стадии проектирования системы изучить все ее основные свойства, в частности эффективность ее функционирования с учетом всех действующих на нее случайных

*) Для этого в [94] служит программа KOLM2, использующая для вычисления $K(u)$ программу SMIRN.

возмущений. В таких случаях прибегают к статистическому моделированию изучаемых явлений. Статистическое моделирование в простейшей форме издавна применялось для решения различных задач. Современная вычислительная техника дает возможность имитировать практически без ограничений сложнейшие явления и процессы. Это привело к созданию и развитию *метода статистического моделирования* как научного метода исследования, позволяющего сочетать теоретические расчеты с имитацией различных экспериментов, а частично и с натурными экспериментами над отдельными элементами исследуемых систем *).

Метод статистического моделирования основан на моделировании (имитации) изучаемого явления на ЭВМ с помощью теоретических зависимостей с непосредственным моделированием влияющих на его течение простейших (первичных) случайных факторов и на статистической обработке получаемых результатов.

Метод статистического моделирования оказался плодотворным и для решения ряда математических задач, не связанных с теорией вероятностей, например для вычисления интегралов, особенно многократных, решения некоторых уравнений и т. п. Применение этого метода к подобным задачам основано на моделировании такого случайного явления, вероятностными характеристиками которого служат искомые величины.

Основой метода статистического моделирования является моделирование случайных величин с заданными распределениями и событий с заданными вероятностями.

4.2. Моделирование случайных величин. Рассмотрим сначала способы моделирования скалярных случайных величин. На основе результатов п. 5.2.3 для моделирования случайной величины X с любой непрерывной функцией распределения $F(x)$ достаточно уметь моделировать случайную величину Y, равномерно распределенную на интервале $[0, 1]$. Получив реализацию y величины Y, найдем соответствующую реализацию x величины X по формуле $x = F^{-1}(y)$.

Для получения реализаций равномерно распределенных случайных величин применяются два принципа: физический и алгоритмический.

---

*) Метод статистического моделирования часто, особенно в иностранной литературе, называют *методом Монте-Карло.*

Физический принцип основан на использовании случайных физических явлений, например потоков частиц, испускаемых радиоактивными веществами, или собственных шумов электронных устройств. Действуя на соответствующие регистрирующие устройства, используемое физическое явление генерирует последовательность независимых случайных величин $V_1, V_2, \ldots$, каждая из которых с вероятностью 0,5 принимает значение 0 или 1. Эту последовательность можно рассматривать как последовательность двоичных знаков числа, представляющего собой значение случайной величины

$$Y = V_1 2^{-1} + V_2 2^{-2} + \ldots + V_k 2^{-k} + \ldots$$

Можно доказать, что при неограниченной длине последовательности величина Y равномерно распределена на интервале $[0, 1]$ *). Реализации этой случайной величины Y вводятся в ЭВМ двоичными числами с тем числом знаков, с которым ЭВМ оперирует.

Для получения реализаций равномерно распределенной случайной величины алгоритмическим путем в ЭВМ вводят произвольное двоичное число, занимающее все разряды или часть разрядов какой-либо ячейки оперативной памяти. Над этим числом выполняют ряд простейших операций по специальной программе. Над полученным числом вновь повторяют те же операции и т. д. Генерируемая таким способом последовательность двоичных чисел, конечно, не будет случайной. Однако в достаточно длинном отрезке этой последовательности все числа с данным числом двоичных знаков будут встречаться практически одинаково часто. Поэтому взятое наугад число из этой последовательности можно считать реализацией случайной величины, равномерно распределенной на интервале $[0, 1]$. Благодаря этому свойству таких последовательностей они называются последовательностями *псевдослучайных чисел*. Существуют много различных алгоритмов и программ формирования псевдослучайных чисел [100] **).

*) См., например, [11, 30]. В действительности вследствие отсутствия строгой независимости величин $V_1, V_2, \ldots$ и из-за того, что вероятности значений 0 и 1 величин V_k не равны 1/2, распределение величины Y оказывается лишь приближенно равномерным.

**) Для генерирования последовательности псевдослучайных чисел, равномерно распределенных на интервале [0,1], в [94] служит программа RANDU.

В отличие от последовательностей, генерируемых физическими источниками случайных явлений, одна и та же последовательность псевдослучайных чисел может неограниченно повторяться. Поэтому одну и ту же реализацию случайной величины можно использовать многократно. Это дает некоторые преимущества при решении ряда практических задач.

Нормально распределенную случайную величину на основании предельной теоремы п. 5.4.1 можно моделировать как сумму независимых случайных величин, распределенных по любым законам. В частности, можно взять сумму независимых равномерно распределенных на интервале [0, 1] случайных величин $Y_1, \ldots, Y_n$. На основании предельной теоремы и результатов примера 3.3 случайная величина

$$X = \frac{\sum_{i=1}^{n} (Y_i - n/2)}{\sqrt{n/12}}$$

при достаточно большом n распределена приблизительно $N(0, 1)$. При этом случайная величина $Z = m_z + X\sqrt{D_z}$ будет распределена приблизительно $N(m_z, D_z)$ *).

Моделирование векторной случайной величины сводится к последовательному моделированию ее координат. При этом первая координата моделируется в соответствии с ее безусловным распределением, а каждая последующая координата — в соответствии с условным распределением при полученных ранее значениях всех предыдущих координат. Для моделирования случайного вектора X, распределенного нормально $N(m_x, K_x)$, можно воспользоваться любым каноническим разложением этого вектора:

$$X = m_x + \sum_{p=1}^{r} V_p x_p,$$

где r — ранг матрицы K_x. Выбрав координатные векторы так, чтобы дисперсии величины V_p были равны 1 (§ 3.4), сведем задачу к моделированию независимых случайных величин $V_1, \ldots, V_r$, распределенных N (0, 1).

*) Для моделирования нормально распределенной случайной величины таким методом в [94] служит программа GAUSS, в которой принято $n = 12$.

4.3. Моделирование событий. Для моделирования события A с вероятностью p достаточно моделировать случайную величину Y, равномерно распределенную в интервале $[0, 1]$. При попадании этой величины в интервал $[0, p]$ считают, что событие A появилось, а при непопадании в интервал $(p, 1]$ — что A не появилось.

Точно так же для моделирования полной группы несовместных событий $A_1, \ldots, A_r$ с вероятностями $p_1, \ldots, p_r$, $p_1 + \ldots + p_r = 1$, достаточно моделировать величину Y, равномерно распределенную в интервале $[0, 1]$, и считать, что появилось A_1 при попадании Y в интервал $[0, p_1]$, появилось A_2 при попадании Y в интервал $(p_1, p_1 + p_2]$, и т. д., появилось A_r при попадании Y в интервал $(p_1 + \ldots + p_{r-1}, 1]$.

4.4. Практические приложения метода. Применение метода статистического моделирования для решения вероятностных задач сводится к многократному непосредственному моделированию на ЭВМ изучаемого явления, включая моделирование тех случайных величин и событий, вероятностные характеристики которых известны, и последующему статистическому оцениванию вероятностных характеристик полученных результатов.

Так, например, при проектировании какой-либо сложной технической системы обычно многократно моделируют функционирование системы и действующие на нее возмущения (входные сигналы), вероятностные характеристики которых предполагаются известными. В результате моделирования получается ряд реализаций (выборка) всех величин, характеризующих работу системы (выходных сигналов). Статистическая обработка полученных данных дает оценки вероятностных характеристик величин, определяющих качество работы системы (как правило, ее точность). Выполнив моделирование для ряда вариантов и значений параметров системы, можно выбрать для дальнейшей разработки и реализации тот вариант, который наилучшим образом удовлетворяет поставленным требованиям.

Большим достоинством метода статистического моделирования является возможность включения в процесс моделирования некоторых реальных элементов системы, в том числе и не поддающихся математическому описанию, например человека или коллектива людей, участвующих в работе системы. Основным недостатком метода статистического моделирования является необходимость

многократного моделирования случайного явления, чтобы получить выборку, по которой необходимые статистические характеристики могут быть оценены с достаточной точностью (*представительную выборку*).

4.5. Точность метода. Чтобы дать представление о зависимости точности метода статистического моделирования от числа опытов, воспользуемся результатами § 6.4, где было показано, что точность оценок

$$\hat{m}_y = \bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k, \qquad \hat{D}_y = \frac{1}{n-1}\sum_{k=1}^{n}(y_k - \bar{y})^2$$

математического ожидания и дисперсии нормально распределенной величины Y характеризуется средними квадратическими отклонениями

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}, \qquad \sigma_{\hat{D}_y} = \sigma_y^2 \sqrt{\frac{2}{n-1}},$$

где σ_y — среднее квадратическое отклонение величины Y *). В приведенной таблице показана зависимость относительной точности оценок математического ожидания и дисперсии от числа опытов n.

Относительная погрешность в %	n								
	10	20	30	40	60	80	100	150	200
$\frac{\sigma_{\bar{y}}}{\sigma_y} \cdot 100$	31,6	22,4	18,2	15,8	12,9	11,2	10	8,2	7,1
$\frac{\sigma_{\hat{D}_y}}{\sigma_y^2} \cdot 100$	47,2	32,4	26,3	22,6	18,4	15,9	14,2	11,6	10,0

Из этой таблицы видно, что математическое ожидание можно оценить со средней квадратической погрешностью 16—19%, а дисперсию—со средней квадратической погрешностью 23—27% при $n = 30-40$. Практически рекомендуется выбирать $n \geqslant 100$, что обеспечит оценивание математического ожидания и дисперсии со средними квадратическими погрешностями 10% и 15% соответственно.

*) Формула для $\sigma_{\bar{y}}$ верна при любом распределении величины Y (пример 3.12).

4.6. Решение вероятностных задач. Простейшей вероятностной задачей, которую можно решить с помощью статистического моделирования, является задача вычисления вероятности попадания векторной случайной величины, особенно многомерной, в данную область B. Непосредственное вычисление интеграла (2.9), как правило, невозможно, а приближенное вычисление путем численного интегрирования требует большого объема вычислений. Моделируя случайную величину X с заданной плотностью $f(x)$, можно за оценку вероятности $P(X \in B)$ принять частоту попадания X в область B. Необходимый для расчетов объем вычислений будет, как правило, меньше, чем при численном интегрировании.

Точно так же можно оценивать моменты различных функций случайной величины X как выборочные средние соответствующих функций, полученные в результате статистического моделирования величины X.

4.7. Вычисление интегралов. Чтобы показать применение метода статистического моделирования для решения математических задач, не имеющих вероятностного содержания, рассмотрим задачу вычисления интеграла

$$I = \int_B g(x)\,dx.$$

Если область B ограничена, то ее можно заключить в некоторый прямоугольный параллелепипед R со сторонами, параллельными осям координат. Представим интеграл I в виде

$$I = \int_B g(x)\,dx = \frac{1}{v(R)} \int_B g(x)\,v(R)\,dx = \frac{1}{v(R)} \int_R g(x)\,\mathbf{1}_B(x)\,v(R)\,dx,$$

где $v(R)$ — объем параллелепипеда R, а $\mathbf{1}_B(x)$ — индикатор области B, т. е. функция, равная 1 при $x \in B$ и 0 при $x \notin B$. Последний интеграл можно рассматривать как математическое ожидание функции $\varphi(X) = g(X)\,\mathbf{1}_B(X)\,v(R)$ случайной величины X, равномерно распределенной в параллелепипеде R. Моделируя эту случайную величину X, можно получить оценку $\hat{I}$ интеграла I как выборочное

среднее случайной величины $\varphi(X)=g(X)\,\mathbf{1}_B(X)\,v(R)$:

$$\hat{I}=\frac{v(R)}{n}\sum_{i=1}^{n} g(x_i)\,\mathbf{1}_B(x_i).$$

При произвольной области B можно взять произвольную плотность $f(x)$, не обращающуюся в нуль нигде в B, например невырожденную нормальную, и представить интеграл I в виде

$$I=\int_B g(x)\,dx=\int_B \frac{g(x)}{f(x)} f(x)\,dx=\int_{-\infty}^{\infty} \frac{g(x)\,\mathbf{1}_B(x)}{f(x)} f(x)\,dx.$$

Последний интеграл можно рассматривать как математическое ожидание функции $\varphi(X)=g(X)\,\mathbf{1}_B(X)/f(X)$ случайной величины X с плотностью $f(x)$. Моделируя эту случайную величину X, можно получить оценку $\hat{I}$ интеграла I как выборочное среднее величины $\varphi(X)=g(X)\,\mathbf{1}_B(X)/f(X)$:

$$\hat{I}=\frac{1}{n}\sum_{i=1}^{n} \frac{g(x_i)\,\mathbf{1}_B(x_i)}{f(x_i)}.$$

Чтобы дать представление о преимуществах метода статистического моделирования перед методами численного интегрирования, предположим, что для вычисления r-кратного интеграла I методами численного интегрирования требуется взять в области B m^r узлов и что для расчета одного значения подынтегральной функции $g(x)$ необходимо sr операций. Тогда для вычисления I методами численного интегрирования потребуется srm^r операций. При применении метода статистического моделирования для вычисления I необходимо srn операций, где n — число опытов. Следовательно, относительный выигрыш машинного времени при применении метода статистического моделирования характеризуется отношением m^r/n. Если считать минимальным необходимым числом узлов по каждой координатной оси $m=5$, то при $n=100$ метод статистического моделирования даст существенную экономию машинного времени уже при $r=4\div 5$. При увеличении кратности интеграла r преимущество метода статистического моделирования растет как показательная функция.

Глава 9

СТАТИСТИЧЕСКИЕ МОДЕЛИ, I

§ 1. Математические модели

1.1. Модели теоретические и статистические. Для применения математических методов необходимо прежде всего установить соотношения между величинами, характеризующими рассматриваемые явления. Каждое такое соотношение представляет собой математическую модель данного явления. Так, например, законы Ньютона в механике и вся построенная на них классическая механика представляют собой совокупность моделей механических явлений. Уравнения Максвелла в физике представляют собой математическую модель электродинамических явлений.

В теории управления изучаются различные *системы*, поведение которых описывается некоторой совокупностью величин, причем одни величины носят характер внешних воздействий на систему и называются ее *входными сигналами*, а другие характеризуют работу системы, определяют ее воздействие на другие системы и называются ее *выходными сигналами*. Зависимость выходных сигналов системы от ее входных сигналов, используемая для изучения системы и ее взаимодействия с другими системами, представляет собой математическую модель этой системы.

Во многих случаях математическую модель можно построить чисто математическим путем на основе известных законов механики, физики и других дисциплин, пользующихся количественными соотношениями. Так, например, различные модели управляемого летательного аппарата можно построить математически, пользуясь законами механики.

Однако существуют и такие системы, для которых принципиально невозможно построить модели чисто математическим путем. Примерами таких систем могут слу-

жить человек или коллектив людей, выполняющий определенные функции, завод, отрасль промышленности, экономика и т. п. Современный научно-технический прогресс и все расширяющееся применение математических методов требуют создания математических моделей и для таких систем. Они необходимы, в частности, для решения современных задач управления.

Для построения моделей сложных систем приходится прибегать к экспериментальному исследованию самих систем или входящих в них подсистем и строить соответствующие модели путем статистической обработки полученных данных.

Модели, полученные на основе статистической обработки результатов экспериментального исследования функционирования систем, будем называть *статистическими моделями*. Методы построения статистических моделей составляют важный раздел современной математической статистики.

Построение математической модели реальной системы или процесса часто называется *идентификацией* этой системы или процесса. Следует заметить, что это название не точно отражает существо данной задачи, так как никакая модель не может быть тождественна соответствующему реальному явлению *).

1.2. Модели детерминированные и стохастические. Из сказанного следует, что задача построения модели явления, процесса или системы состоит в нахождении соотношений между величинами, описывающими течение данного явления, процесса или функционирование данной системы. Если эти соотношения позволяют по данным значениям одних величин однозначно определить значения других, то описываемая ими модель называется *детерминированной*. Если же эти соотношения по данным значениям одних величин определяют другие как случайные величины, то описываемая ими модель называется *стохастической*.

Как теоретические модели, выводимые математически из первичных законов, так и статистические модели, получаемые на основе статистической обработки результатов наблюдений, могут быть детерминированными или стохастическими. Так, например, зависимость $Y = \varphi(X)$,

---

*) Английские слова «identity» и «identification» означают «тождество» и «отождествление».

устанавливаемая по результатам наблюдений случайных величин X и Y методом наименьших квадратов, представляет собой детерминированную модель. Если же учесть наблюдаемые в результате опытов случайные отклонения экспериментальных точек от кривой $y=\varphi(x)$ и написать зависимость Y от X в виде

$$Y=\varphi(X)+Z,$$

где Z — некоторая случайная величина, то получится стохастическая модель. При этом величины X и Y могут быть как скалярными, так и векторными. Функция $\varphi(x)$ может быть как линейной комбинацией данных функций, так и данной нелинейной функцией, параметры которой определяются методом наименьших квадратов. Число и вид функций в линейной комбинации, а также вид нелинейной функции $\varphi(x)$ и число определяющих ее параметров могут быть любыми.

Приведенный простой пример показывает, что одному и тому же явлению могут соответствовать различные модели. Проблема построения модели системы включает и выбор подходящей формы этой модели, а также разумной степени ее сложности, совместимых с имеющимися экспериментальными данными.

1.3. Роль математических моделей. Математические модели являются мощным инструментом познания реального мира. Они применяются для различных расчетов, проектирования систем, управления процессами и системами, прогнозирования разных явлений и т. д. Математически модели дают возможность открывать и предсказывать новые научные факты и закономерности.

Применение математических моделей в основном сводится к определению значений некоторых величин по известным значениям других величин. Значения последних могут быть или получены в результате наблюдений, или задаваться из каких-либо соображений. Так, например, измеряя какую-либо величину, мы наблюдаем только результаты измерений, по которым требуется определить значение самой измеряемой величины. При прогнозировании погоды по измеренным значениям параметров состояния атмосферы в одних точках пространства в один период времени предсказываются их значения в других точках и в другой период времени. При применении математических методов в медицинской практике по результатам обследования пациента должна решаться задача

распознавания болезни, которой он страдает, и определения соответствующих методов лечения. Аналогичные задачи распознавания возникают во многих других областях.

Современное состояние и развитие вычислительной техники дают возможность автоматизировать некоторые виды умственной деятельности человека, т. е. приступить к решению проблемы создания искусственного интеллекта. Это особенно важно для разработки автоматических устройств — роботов — для выполнения работ в таких условиях, в которых человек не может существовать. Такие роботы должны согласовывать свои действия с изменяющимися условиями и вновь возникающими обстоятельствами и, следовательно, распознавать ситуации и в соответствии с этим принимать решения. Не менее важны для современного научно-технического прогресса задачи создания автоматических систем, читающих и анализирующих рукописные тексты, а также систем, анализирующих человеческую речь. Такие системы должны распознавать рукописные буквы, цифры и другие изображения или соответственно звуки человеческой речи (фонемы) и определять их смысловое содержание. Объединив алгоритм распознавания человеческой речи с алгоритмами перевода с одного языка на другой, можно будет создавать автоматические системы для синхронного перевода речи на несколько языков. Наконец, при использовании ЭВМ для обучения людей также необходимы алгоритмы автоматического распознавания характера ошибок, допускаемых обучаемыми, и определения уровня их знаний. Человек в своей практической деятельности повседневно решает различные задачи распознавания. Для создания автоматических распознающих систем необходимы соответствующие модели процессов распознавания.

Все расширяющееся проникновение математических методов в сферу управления ставит задачу создания моделей и для процессов принятия решений. Руководитель любой организации по полученной информации (которую всегда можно представить в виде совокупности значений некоторых величин) должен принять решение (которое тоже можно представить в виде совокупности значений определенных величин). Современная вычислительная техника может помогать руководителям принимать наиболее научно обоснованные решения и тем самым повышать эффективность управления, но для этого необходимы соот-

ветствующие модели процессов обоснования и принятия решений.

Перейдем теперь к изучению методов построения статистических моделей. При этом мы ограничимся только такими моделями, которые описываются соотношениями между конечномерными векторами. К таким моделям относятся, в частности, все дискретные системы, входные и выходные сигналы которых действуют только в определенные дискретные моменты времени. Для построения более сложных статистических моделей, в частности моделей, входные и выходные сигналы которых представляют собой функции непрерывно изменяющихся аргументов, необходимо применение теории случайных функций — раздела теории вероятностей, выходящего за рамки этой книги *).

§ 2. Регрессионные модели

2.1. Регрессия как оценка зависимости случайной величины от другой величины. Рассмотрим задачу определения значения случайной величины по данным значениям другой величины. Пусть Y — случайная величина, значения которой требуется определить, x — известная величина, которая может представлять собой наблюденное значение некоторой случайной величины X или заданное значение некоторой переменной. Как x, так и Y могут быть скалярными или векторными величинами. Распределение случайной величины Y в общем случае зависит от x.

С точки зрения математической статистики поставленная задача представляет собой задачу оценивания значения ненаблюдаемой случайной величины Y по данному значению величины x (в частности, по полученному в результате наблюдения значению x случайной величины X).

Пусть $\hat{y} = \hat{y}(x)$ — оценка значения величины Y при данном x. Ошибка этой оценки $\hat{y} - Y$ представляет собой случайную величину. В соответствии с п. 6.2.1 точность оценки $\hat{y}$ целесообразно характеризовать средним квадратом ошибки при данном значении x:

$$\varepsilon(x) = M[\,|\hat{y}(x) - Y|^2 \,|\, x]. \qquad (1)$$

*) Эти вопросы будут рассмотрены в книге «*Случайные функции*», которую автор готовит к печати.

Рассмотрим сначала случай скалярной величины Y. Из формулы (3.31) следует, что средний квадрат ошибки оценки $\hat{y}$ будет минимальным, если принять за $\hat{y}$ математическое ожидание случайной величины Y при данном x:

$$\hat{y}(x) = m_y(x) = M[Y \mid x]. \qquad (2)$$

Эта формула определяет наилучшую оценку величины Y при данном x.

В случае векторной величины $Y = [Y_1 \dots Y_m]^T$

$$\varepsilon(x) = M[|\hat{y} - Y|^2 \mid x] = \sum_{p=1}^{m} M[|\hat{y}_p - Y_p|^2 \mid x]. \qquad (3)$$

Отсюда видно, что величина $\varepsilon(x)$ минимальна тогда и только тогда, когда каждое слагаемое в (3) имеет минимальное значение. Следовательно, формула (2) определяет наилучшую с точки зрения минимума величины ε оценку при каждом x и для векторной случайной величины Y.

Зависимость оценки $\hat{y}$ величины Y от x, определяемая формулой (2), представляет собой регрессию Y на x (п. 4.3.2). Таким образом, *оптимальной оценкой зависимости Y от x служит регрессия Y на x*. В частности, *оптимальным прогнозом величины Y по данному значению x будет прогноз по регрессии.*

Модель, определяемая регрессией Y на x, называется *регрессионной моделью*. Регрессионные модели широко применяются для описания функционирования различных систем.

2.2. Средние квадратические регрессии различных классов. Оценивание по регрессии возможно только в том случае, когда регрессия известна. Если регрессия не известна или определяемая ею зависимость слишком сложна для практической реализации, то приходится искать оценку зависимости Y от x в некотором ограниченном классе функций. При этом уже не удается обеспечить минимум среднего квадрата ошибки $\varepsilon(x)$ при каждом значении x. Поэтому обычно ищут оценку зависимости Y от x в данном классе функций из условия минимума среднего значения величины $\varepsilon(x)$ в интересующей нас области изменения x.

Предположим сначала, что x представляет собой значение некоторой случайной величины X. Тогда естественно рассматривать средний квадрат ошибки как функцию случайной величины X и за ее среднее значение

в области изменения x принять ее математическое ожидание. Пусть $\hat{y}(x)$ — искомая оценка зависимости Y от x в некотором классе функций Ψ. Рассматривая эту оценку как функцию $\hat{y}(X)$ случайной величины X, получим на основании (1) значение среднего квадрата ошибки оценки как функцию случайной величины X:

$$\varepsilon(X) = M[|\hat{y}(X) - Y|^2 \mid X].$$

Взяв математическое ожидание этой величины и пользуясь формулой полного математического ожидания (4.34), получим среднее значение величины ε:

$$\varepsilon = M|\hat{y}(X) - Y|^2. \tag{4}$$

Предположим теперь, что x — неслучайная переменная, изменяющаяся в некоторой области B, и за среднее значение величины $\varepsilon(x)$ принимается

$$\varepsilon = \int_B p(x)\,\varepsilon(x)\,dx,$$

где $p(x)$ — некоторая функция веса, обладающая свойством (как и всякая функция веса при определении взвешенного среднего)

$$\int_B p(x)\,dx = 1.$$

В частном случае, когда всем значениям x в области B придается один и тот же вес, $p(x) = 1/v(B)$, где $v(B)$ — объем области B. Очевидно, что этот случай сводится к предыдущему, если считать x значением случайной величины X, плотность которой равна $p(x)$ при $x \in B$ и 0 при $x \notin B$. Следовательно, формулу (4) можно считать общей формулой для определения среднего значения величины ε в интересующей нас области изменения x.

Таким образом, наилучшей оценкой зависимости Y от x в заданном классе функций Ψ является функция $\hat{y}(x)$, реализующая минимум безусловного математического ожидания квадрата модуля ошибки.

Оценка $\hat{y}(X)$ случайной величины Y, принадлежащая определенному классу функций Ψ, для которой математическое ожидание квадрата модуля ошибки (средний квадрат ошибки) достигает минимума, называется *средней квадратической регрессией Y на x класса Ψ*. В частности, линейная функция, реализующая минимум среднего квад-

рата ошибки ε в классе линейных функций, называется *линейной средней квадратической регрессией Y на x*.

2.3. Оптимальные оценки. Покажем, что *средняя квадратическая регрессия класса Ψ представляет собой наилучшее среднее квадратическое приближение регрессии $m_y(x)=M[Y|x]$ в этом классе функций Ψ.*

▶ Заметим, что на основании (3.31)

$$M[|\hat{y}(X)-Y|^2|X]=\sum_{p=1}^{m} M[|\hat{y}_p(X)-Y_p|^2|X]=$$
$$=\sum_{p=1}^{m}\{M[|m_p(X)-Y_p|^2|X]+|\hat{y}_p(X)-m_y(X)|^2\}=$$
$$=M[|m_y(X)-Y|^2|X]+|\hat{y}(X)-m_y(X)|^2.$$

Взяв математическое ожидание этой случайной величины, на основании формулы полного математического ожидания (4.34) получаем

$$\varepsilon=M|\hat{y}(X)-Y|^2=M|m_y(X)-Y|^2+M|\hat{y}(X)-m_y(X)|^2. \tag{5}$$

Так как первое слагаемое в правой части не зависит от функции $\hat{y}(X)$, то ε достигает минимума тогда и только тогда, когда минимально второе слагаемое. ◀

Оценку $\hat{y}(x)$ регрессии $m_y(x)=M[Y|x]$, минимизирующую средний квадрат ошибки ε в классе функций Ψ, будем называть *оптимальной оценкой в классе Ψ*.

Из (5) видно, что если регрессия $m_y(x)$ содержится в классе функций Ψ, в котором ищется оптимальная оценка, в частности, если Ψ совпадает с множеством всех функций переменной x, то оптимальной оценкой регрессии служит сама регрессия.

Регрессия является единственной оценкой, реализующей минимум условного математического ожидания квадрата модуля ошибки при каждом данном значении x.

2.4. Необходимое и достаточное условие оптимальности. Покажем, что $\hat{y}(x)$ *представляет собой оптимальную оценку в классе функций Ψ, если для любой функции $\psi(x)\in\Psi$*

$$M\psi(X)^*[\hat{y}(X)-Y]=\operatorname{tr}M[\hat{y}(X)-Y]\psi(X)^*=0\ ^{*}). \tag{6}$$

---

*) Как всегда, звездочка означает транспонирование матрицы с заменой всех ее элементов соответствующими сопряженными числами. Первое равенство (6) верно в силу соотношения $ab=\operatorname{tr} ba$, справедливого для любых матриц-столбцов (векторов) a и b.

▶ Пусть $\psi(x)$ — произвольная функция класса Ψ, $\hat{y}(x)$ — оптимальная оценка Y в классе Ψ. Если принять за оценку Y функцию $\psi(X)$, то будем иметь

$$\varepsilon = M|\psi(X)-Y|^2 = M|\hat{y}(X)-Y+\psi(X)-\hat{y}(X)|^2 = \\ = M|\hat{y}(X)-Y|^2 + M|\psi(X)-\hat{y}(X)|^2 + \\ + M[\psi(X)^* - \hat{y}(X)^*][\hat{y}(X)-Y] + \\ + M[\hat{y}(X)^* - Y^*][\psi(X)-\hat{y}(X)]. \quad (7)$$

Если условие (6) выполнено для любой функции $\psi(x) \in \Psi$, то оно выполнено и для $\psi(x) = \hat{y}(x)$. В этом случае

$$M[\psi(X)^* - \hat{y}(X)^*][\hat{y}(X)-Y] = 0,$$

$$M[\hat{y}(X)^* - Y^*][\psi(X)-\hat{y}(X)] = \\ = \overline{M[\psi(X)^* - \hat{y}(X)^*][\hat{y}(X)-Y]} = 0$$

и (7) принимает вид

$$\varepsilon = M|\psi(X)-Y|^2 = M|\hat{y}(X)-Y|^2 + M|\psi(X)-\hat{y}(X)|^2.$$

Отсюда видно, что ε достигает минимума тогда и только тогда, когда $\psi(x) \equiv \hat{y}(x)$. Таким образом, условие (6) достаточно для оптимальности оценки $\hat{y}(x)$. ◀

Докажем, что *если класс функций* Ψ *представляет собой линейное пространство* *), *то условие* (6) *необходимо для оптимальности оценки* $\hat{y}(x) \in \Psi$.

---

*) Множество L называется *линейным* или *векторным пространством*, если оно удовлетворяет условиям:

1) для любых x, $y \in L$ однозначно определен элемент $z \in L$, называемый суммой x и y, $z = x + y$;

2) для любого $x \in L$ и числа α однозначно определен элемент $u \in L$, называемый произведением x на число α, $u = \alpha x$;

3) в L существует нулевой элемент 0 такой, что $x + 0 = x$, $0 \cdot x = 0$ для всех $x \in L$;

4) операции сложения и умножения на число обладают свойствами:

$x + y = y + x$ (коммутативность);

$(x+y)+z = x+(y+z)$, $\alpha(\beta x) = (\alpha\beta)x$ (ассоциативность);

$\alpha(x+y) = \alpha x + \alpha y$, $(\alpha+\beta)x = \alpha x + \beta x$ (дистрибутивность).

В зависимости от того, определена операция умножения на число для всех комплексных чисел или только для действительных, L называется *комплексным линейным пространством* или, соответственно, *действительным линейным пространством* [42].

▶ Предположив, что существует функция $\psi_0(x) \in \Psi$, для которой

$$M\psi_0(X)^* [\hat{y}(X) - Y] = \alpha \neq 0,$$

определим функцию

$$\psi(x) = \hat{y}(x) - \frac{\alpha\psi_0(x)}{M\,|\,\psi_0(X)\,|^2}.$$

Так как Ψ — линейное пространство, то $\psi(x) \in \Psi$. Вычислим средний квадрат ошибки ε для оценки $\psi(x)$. В данном случае

$$M\,|\,\psi(X) - \hat{y}(X)\,|^2 = \frac{|\,\alpha\,|^2}{M\,|\,\psi_0(X)\,|^2},$$

$$M[\psi(X)^* - \hat{y}(X)^*][\hat{y}(X) - Y] =$$
$$= -\frac{\bar{\alpha}}{M\,|\,\psi_0(X)\,|^2} M\psi_0(X)^* [\hat{y}(X) - Y] = -\frac{|\,\alpha\,|^2}{M\,|\,\psi_0(X)\,|^2},$$

$$M[\hat{y}(X)^* - Y^*][\psi(X) - \hat{y}(X)] =$$
$$= \overline{M[\psi(X)^* - \hat{y}(X)^*][\hat{y}(X) - Y]} = -\frac{|\,\alpha\,|^2}{M\,|\,\psi_0(X)\,|^2}.$$

Подставив эти выражения в (7), получим

$$\varepsilon = M\,|\,\psi(X) - Y\,|^2 = M\,|\,\hat{y}(X) - Y\,|^2 - \frac{|\,\alpha\,|^2}{M\,|\,\psi_0(X)\,|^2}.$$

Отсюда следует, что оценка $\psi(x)$ лучше оценки $\hat{y}(x)$. Таким образом, если существует функция $\psi_0(x) \in \Psi$, для которой условие (6) не выполнено, то оценка $\hat{y}(x)$ не может быть оптимальной. Следовательно, выполнение условия (6) для всех $\psi(x) \in \Psi$ необходимо. ◀

2.5. Линейные регрессионные модели. Найдем оптимальную оценку $\hat{y}(x)$ величины Y в классе линейных функций, т. е. линейную среднюю квадратическую регрессию Y на X.

Положив в (6) $\hat{y}(X) = gX$, $\psi(X) = hX$, получим

$$\mathrm{tr}\, M(gX - Y)X^*h^* = 0,$$

или, в силу того, что $MXX^* = \Gamma_x$, $MYX^* = \Gamma_{yx}$ (§ 3.3),

$$\mathrm{tr}\,(g\Gamma_x - \Gamma_{yx})\,h^* = 0. \qquad (8)$$

Это равенство справедливо для любой матрицы h тогда и только тогда, когда

$$g\Gamma_x = \Gamma_{yx}. \qquad (9)$$

Действительно, если какие-нибудь элементы матрицы $\alpha = g\Gamma_x - \Gamma_{yx}$ отличны от нуля, то, положив $h = \alpha$, получим $\operatorname{tr}(g\Gamma_x - \Gamma_{yx}) h^* = \operatorname{tr} \alpha\alpha^* \neq 0$, если хотя бы один элемент матрицы α отличен от нуля *).

Так как класс линейных функций представляет собой линейное пространство, то (8), а следовательно и (9), необходимо и достаточно для оптимальности оценки $\hat{y}(x) = gx$. Таким образом, матрица g, соответствующая оптимальной линейной оценке, определяется уравнением (9).

Заметим, что уравнение (9) можно представить в виде

$$M(\hat{Y} - Y) X^* = 0, \tag{10}$$

где $\hat{Y} = gX$. Из (8) следует также, что для любого разбиения вектора X на блоки, $X^T = [X_1^T, \ldots, X_N^T]$, оптимальная оценка $\hat{Y}$ удовлетворяет также условиям

$$M(\hat{Y} - Y) X_k^* = 0 \qquad (k = 1, \ldots, N). \tag{11}$$

Действительно, разбив матрицу h на соответствующие блоки $h = [h_1 \ldots h_N]$ и положив $h_i = 0$ при $i \neq k$, выведем (11) из (8) совершенно так же, как было получено (9).

Оценка $\hat{y}(x) = gx$, определяемая уравнением (9), в общем случае будет смещенной, $M\hat{y}(X) = M(gX) \neq MY$. Для устранения смещения достаточно добавить к gx соответствующее постоянное слагаемое.

▶ Положив в (6) $\hat{y}(X) = gX + a$, $\psi(X) = hX + b$, в силу соотношений

$$\Gamma_x = K_x + m_x m_x^*, \qquad \Gamma_{yx} = K_{yx} + m_y m_x^* \tag{12}$$

получаем

$$\operatorname{tr}[(gK_x - K_{yx}) h^* + (gm_x - m_y + a)(m_x^* h^* + b^*)] = 0. \tag{13}$$

Это условие должно выполняться для любых матрицы h и вектора b. Положив, в частности, $b = -hm_x$, будем иметь

$$\operatorname{tr}(gK_x - K_{yx}) h^* = 0.$$

---

*) След матрицы $\alpha\alpha^*$ равен сумме квадратов модулей всех элементов матрицы α:

$$\operatorname{tr} \alpha\alpha^* = \sum_{p=1}^{m} \left(\sum_{q=1}^{m} \alpha_{pq} \bar{\alpha}_{pq} \right) = \sum_{p,\, q=1}^{m} |\alpha_{pq}|^2.$$

Отсюда совершенно так же, как в предыдущем случае, получаем уравнение для матрицы g:

$$gK_x = K_{yx}. \tag{14}$$

При этом условие (13) будет выполняться для любых матрицы h и вектора b тогда и только тогда, когда

$$a = m_y - gm_x. \tag{15}$$

Таким образом, оптимальная линейная оценка со сдвигом $\hat{y} = gx + a$ (линейная средняя квадратическая регрессия) определяется уравнением (14) и формулой (15), которые представляют собой необходимые и достаточные условия оптимальности. ◀

Если регрессия Y на X линейна, то на основании сделанного выше замечания функция $\hat{y}(x) = gx + a$, где g и a определяется уравнением (14) и формулой (15), совпадает с регрессией.

Заметим, что линейную функцию со сдвигом $gx + a$ можно рассматривать как частный случай линейной функции gx. Действительно, добавив к вектору x еще одну координату, равную 1, т. е. положив $x' = [x^T \ 1]^T$, и вводя матрицу $g' = [g \ a]$, получим $gx + a = g'x'$. Соответственно уравнение (14) и формула (15) получаются из (9) в частном случае, когда одна из координат вектора x равна 1.

В частном случае, когда векторы X и Y действительны и их совместное распределение нормально, регрессии всегда линейны (§ 4.4). Поэтому уравнение (14) и формула (15) определяют регрессию проекции Y нормально распределенного случайного вектора $[X^T \ Y^T]^T$ на его проекцию X на дополнительное подпространство. Формула (4.55) и уравнение (4.57), определяющие регрессию одной координаты нормально распределенного случайного вектора на все остальные его координаты, вытекают из полученных результатов как частные случаи.

Пример 1. Пусть $f(x, y)$ — совместная плотность случайных векторов X и Y, m_x, K_x — математическое ожидание и ковариационная матрица вектора X, причем определитель $|K_x|$ отличен от нуля. В этом случае определитель матрицы $\Gamma_x = K_x + m_x m_x^T$ строго положителен и уравнение (9) имеет единственное решение

$$g = \Gamma_{yx}\Gamma_x^{-1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yx^T\Gamma_x^{-1} f(x, y)\,dx\,dy = \int_{-\infty}^{\infty} m_y(x)\, x^T\Gamma_x^{-1} f_1(x)\,dx,$$

где $f_1(x)$ — плотность вектора X. Эта формула вместе с приближенной формулой $m_y(x) \approx gx$ дает *статистическую линеаризацию* регрессии $m_y(x)$ по Бутону *).

Пример 2. В условиях примера 1 уравнение (14) имеет единственное решение

$$g = K_{yx}K_x^{-1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - m_x)(x - m_x)^T K_x^{-1} f(x, y)\, dx\, dy =$$

$$= \int_{-\infty}^{\infty} [m_y(x) - m_x](x - m_x)^T K_x^{-1} f_1(x)\, dx.$$

Эта формула вместе с приближенной формулой $m_y(x) \approx m_y + g(x - m_x)$ дает *статистическую линеаризацию* регрессии $m_y(x)$ по Казакову. Таким образом, в основе статистической линеаризации лежат линейные регрессионные модели.

2.6. Решение уравнений, определяющих линейную регрессию. Для решения уравнения (14) можно использовать различные вычислительные методы и соответствующие машинные программы [1, 94, 107]. Одним из наиболее удобных методов решения уравнения (14) является метод, основанный на каноническом разложении случайного вектора X. Пусть

$$X = m_x + \sum_{p=1}^{r} V_p x_p \tag{16}$$

— какое-нибудь каноническое разложение вектора X, $f_1, \ldots, f_r$ — векторы, удовлетворяющие вместе с координатными векторами $x_1, \ldots, x_r$ условию биортогональности $f_p^T x_q = x_q^T f_p = \delta_{pq}$ $(p, q = 1, \ldots, r)$. При этом в силу (3.71) и (3.69) $x_p = D_p^{-1} K_x f_p$ и $V_p = f_p^T X^0 = X^{0T} f_p$, где D_p — дисперсия величины V_p.

▶ Будем искать решение уравнения (14) в виде

$$g = \sum_{p=1}^{r} c_p f_p^T. \tag{17}$$

---

*) Бутон [12] и Казаков [31] независимо друга от друга разработали метод статистической линеаризации для приближенного исследования точности нелинейных систем. Они рассматривали случай скалярных X и $Y = \varphi(X)$ при заданной функции $\varphi(x)$. Впоследствии Соммервил и Атертон [101], Казаков [32] и другие распространили этот метод на векторные X и Y (см. также сбзор [95]).

Для определения неизвестных коэффициентов $c_1, \ldots, c_r$ подставим выражение (17) в (14):

$$\sum_{p=1}^{r} c_p f_p^T K_x = K_{yx}.$$

Учитывая, что $f_p^T K_x = D_p x_p^*$, получим

$$\sum_{p=1}^{r} c_p D_p x_p^* = K_{yx}.$$

Умножив это уравнение справа на $\bar{f}_q$ и заметив, что $x_p^* \bar{f}_q = \overline{x_p^T f_q} = \delta_{pq}$, получим

$$c_q D_q = K_{yx} \bar{f}_q.$$

Отсюда находим

$$c_p = \frac{1}{D_p} K_{yx} \bar{f}_p \qquad (p = 1, \ldots, r).$$

Подставив это выражение в (17), будем иметь

$$g = \sum_{p=1}^{r} \frac{1}{D_p} K_{yx} \bar{f}_p f_p^T. \tag{18}$$

Остается убедиться в том, что определяемая этой формулой матрица g удовлетворяет уравнению (14). Для этого заметим, что в силу (16)

$$K_{yx} = MY^0 X^{0*} = \sum_{p=1}^{r} MY^0 \bar{V}_p x_p^*.$$

Так как $V_p = f_p^T X^0 = X^{0T} f_p$, то

$$MY^0 \bar{V}_p = MY^0 X^{0*} \bar{f}_p = K_{yx} \bar{f}_p,$$

$$K_{yx} = \sum_{p=1}^{r} K_{yx} \bar{f}_p x_p^* = \sum_{p=1}^{r} \frac{1}{D_p} K_{yx} \bar{f}_p f_p^T K_x.$$

То же получается, если (18) умножить справа на K_x:

$$gK_x = \sum_{p=1}^{r} \frac{1}{D_p} K_{yx} \bar{f}_p f_p^T K_x.$$

Следовательно, матрица g, определяемая формулой (18), удовлетворяет уравнению (14). ◀

Формула (18) дает решение уравнения (14) как в том случае, когда определитель матрицы K_x отличен от нуля,

так и в том случае, когда он равен нулю. Во втором случае общее решение уравнения (14) получится прибавлением к правой части формулы (18) произвольной матрицы h, удовлетворяющей уравнению $hK_x=0$.

▶ Для решения уравнения (9) представим его с помощью соотношений (12) в виде

$$gK_x = K_{yx} + \lambda m_x^*, \tag{19}$$

где

$$\lambda = m_y - gm_x. \tag{20}$$

Так как уравнение (19) линейное, то его решение имеет вид

$$g = g_0 + \lambda g_1, \tag{21}$$

где g_0 и g_1 — решения уравнений

$$g_0 K_x = K_{yx}, \quad g_1 K_x = m_x^*. \tag{22}$$

Решение первого из этих уравнений дается формулой (18). Чтобы решить второе, представим ковариационную матрицу K_x каноническим разложением (3.75):

$$K_x = \sum_{p=1}^{r} D_p x_p x_p^*.$$

Тогда второе уравнение (22) примет вид

$$\sum_{p=1}^{r} D_p g x_p x_p^* = m_x^*.$$

Очевидно, что это уравнение имеет решение только в том случае, когда математическое ожидание m_x представимо разложением по координатным векторам:

$$m_x = \sum_{p=1}^{r} (f_p^T m_x)\, x_p = \sum_{p=1}^{r} (m_x^T f_p)\, x_p. \tag{23}$$

А это будет или когда определитель матрицы K_x отличен от нуля, или когда вектор m_x принадлежит подпространству, на котором сосредоточено распределение центрированного вектора X^0. Если ни одно из этих условий не выполнено, то второе уравнение (22) не имеет решения. Предположив, что m_x представимо разложением (23), совершенно так же, как и раньше, найдем решение второго уравнения (22) в виде

$$g_1 = \sum_{p=1}^{r} \frac{1}{D_p} m_x^* \bar{f}_p f_p^T. \tag{24}$$

Определив g_0 и g_1 и подставив выражение (21) в (20), найдем λ:

$$\lambda = \frac{m_y - g_0 m_x}{1 - g_1 m_x}. \tag{25}$$

Формулы (18), (24), (25) и (21) дают решение уравнения (9).

Если ранг ρ матрицы K_x меньше размерности r вектора X, то согласно § 3.4 $x_{\rho+1} = \ldots = x_r = 0$, $D_{\rho+1} = \ldots$ $\ldots = D_r = 0$ и суммирование по p в (16), (17), (18), (23) и (24) производится от 1 до ρ.

Если математическое ожидание m_x не может быть представлено разложением (23), то уравнение (19) имеет решение, для которого $\lambda = 0$, и, следовательно, $gm_x = m_y$. Чтобы получить это решение, положим (в этом случае $\rho < r$ и $x_{\rho+1} = \ldots = x_r = 0$)

$$s = m_x - \sum_{p=1}^{\rho} (f_p^T m_x) x_p$$

и определим матрицу-строку

$$h = \frac{s^*}{|s|^2} + \sum_{p=1}^{r} \gamma_p f_p^T$$

так, чтобы было

$$hx_q = 0 \qquad (q = 1, \ldots, \rho).$$

Для этого в силу условия $f_p^T x_q = \delta_{pq}$ достаточно взять $\gamma_q = -s^* x_q / |s|^2$ $(q = 1, \ldots, \rho)$. Очевидно, что матрица h удовлетворяет однородному уравнению

$$hK_x = 0.$$

Чтобы убедиться в этом, достаточно воспользоваться каноническим разложением матрицы K_x. Тогда получим

$$hK_x = h \sum_{p=1}^{\rho} D_p x_p x_p^* = \sum_{p=1}^{\rho} D_p (hx_p) x_p^* = 0.$$

В силу этого матрица

$$g = g_0 + \mu h$$

при любой матрице-столбце μ удовлетворяет уравнению

$$gK_x = K_{yx},$$

т. е. уравнению (19) при $\lambda = 0$. Остается выбрать μ так, чтобы было $\lambda = 0$. Подставив найденное решение g в (20),

находим

$$\lambda = m_y - gm_x = m_y - g_0 m_x - \mu h m_x = m_y - g_0 m_x - \mu,$$

так как $hm_x = hs = s^* s / |s|^2 = 1$. Положив $\mu = m_y - g_0 m_x$, будем иметь $\lambda = 0$. При этом формула

$$g = g_0 + (m_y - g_0 m_x) h \qquad (26)$$

определяет решение уравнения (9). В этом можно убедиться и непосредственной подстановкой выражения (26) в (9) с учетом первого уравнения (22) и равенств $hK_x = 0$, $hm_x = 1$. ◀

2.7. Детерминированные и стохастические регрессионные модели. Регрессия $m_y(x)$ и ее оптимальная оценка $\hat{y}(x)$ представляют собой детерминированные регрессионные модели. Для получения стохастической регрессионной модели достаточно представить величину Y в виде $Y = m_y(x) + Z$ или $Y = \hat{y}(x) + Z_1$, где Z и Z_1 — некоторые случайные величины.

Из (9), (14) и (15) видно, что для нахождения детерминированной линейной регрессионной модели достаточно знать математические ожидания m_x, m_y векторов X и Y и ковариационные матрицы K_x, K_{yx}. Для нахождения стохастической линейной регрессионной модели необходимо еще знать распределение случайной величины Y при любом x или, по крайней мере, ее математическое ожидание $m_y(x)$ (регрессию) и ковариационную матрицу $K_y(x)$ (совпадающую с ковариационной матрицей $K_z(x)$ величины Z).

Более общая задача наилучшего приближения регрессии конечной линейной комбинацией заданных функций $\varphi_1(x), \ldots, \varphi_N(x)$ сводится к задаче наилучшего линейного приближения регрессии, так как любая линейная комбинация функций $\varphi_1(x), \ldots, \varphi_N(x)$ представляет собой линейную функцию переменных $z_1 = \varphi_1(x), \ldots, z_N = \varphi_N(x)$.

Пример 3. Построить оптимальную полиномиальную регрессионную модель в случае скалярных X и Y.

Полагая $z_1 = x$, $z_2 = x^2$, ..., $z_N = x^N$, $z = [1 \; z_1 \ldots z_N]^T$, $g = [a \, g_1 \ldots g_N]$, сведем задачу к нахождению линейной модели $\hat{y}(x) = gz$. Уравнение (9) в этом случае принимает вид

$$[a \; g_1 \ldots g_N] \begin{bmatrix} 1 & \alpha_1 & \ldots & \alpha_N \\ \alpha_1 & \alpha_2 & \ldots & \alpha_{N+1} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_N & \alpha_{N+1} & \ldots & \alpha_{2N} \end{bmatrix} = [m_y \; \gamma_1 \ldots \gamma_N],$$

где $\alpha_p = MX^p$ — p-й момент величины X, а $\gamma_p = MYX^p$ — $(p+1)$-й момент случайного вектора $[X\ Y]^T$, содержащий Y в первой степени. Полученное уравнение определяет полиномиальную среднюю квадратическую регрессию Y на X степени N, которая и служит оптимальной полиномиальной регрессионной моделью.

Аналогично строятся оптимальные полиномиальные регрессионные модели в случае векторных X и Y.

Для решения уравнения, определяющего полиномиальную регрессионную модель, применяются те же вычислительные методы, что и для линейных моделей. В частности, можно пользоваться каноническими разложениями. В рассмотренном случае скалярной величины X для этого достаточно найти методом § 3.4 каноническое разложение случайного вектора $Z = [1\ X \ldots X^N]^T$.

§ 3. Оценивание регрессий

3.1. Оценка матрицы коэффициентов линейной регрессии. Если математические ожидания m_x, m_y и ковариационные матрицы K_x, K_{yx} не известны, то для нахождения линейной регрессионной модели приходится оценивать по результатам опытов матрицу g, а в случае стохастической линейной модели и статистические характеристики величины Y при каждом данном x.

Как всегда в задачах математической статистики, будем считать все величины действительными.

Предположим, что в результате n независимых опытов получены пары значений $(x_1, y_1), \ldots, (x_n, y_n)$ величин x и Y. Для нахождения линейной оценки регрессии $\hat{y}(x) = \hat{g}x$ естественно применить метод наименьших квадратов (п. 1.2.8). Тогда матрица $\hat{g}$ определится из условия минимума величины

$$\varepsilon = \sum_{k=1}^{n} | y_k - \hat{g}x_k |^2.$$

Это выражение отличается от (4) при $\hat{y}(x) = \hat{g}x$ только тем, что вместо знака математического ожидания стоит знак суммы. Поэтому уравнение для матрицы $\hat{g}$ получается заменой в (9) математических ожиданий соответствующими суммами:

$$\hat{g} \sum_{k=1}^{n} x_k x_k^T = \sum_{k=1}^{n} y_k x_k^T. \tag{27}$$

Вводя матрицы

$$A = \sum_{k=1}^{n} x_k x_k^T, \quad B = \sum_{k=1}^{n} y_k x_k^T, \tag{28}$$

перепишем (27) в виде $\hat{g}A=B$. Это уравнение имеет единственное решение тогда и только тогда, когда матрица A неособенная. Этого всегда можно достичь выбором $x_1, \ldots$ $\ldots, x_n$ или изменением числа опытов. В этом случае решение уравнения (27) дается формулой

$$\hat{g}=BA^{-1}=\sum_{k=1}^{n} y_k x_k^T A^{-1}.$$

3.2. Статистические свойства оценки. Изучим статистические свойства найденной оценки матрицы g.

▶ Рассмотрим сначала случай, когда x — неслучайная переменная. В этом случае $\hat{g}$ представляет собой полученную в результате опытов реализацию случайной величины

$$\hat{G}=\sum_{k=1}^{n} Y_k x_k^T A^{-1}, \tag{29}$$

где $Y_1, \ldots, Y_n$ — случайные значения величины Y при n опытах. Из (29) следует, что

$$M\hat{G}=\sum_{k=1}^{n} m_y(x_k)\, x_k^T A^{-1}. \tag{30}$$

В частном случае линейной регрессии $m_y(x)=gx$ формула (30) принимает вследствие (28) вид

$$M\hat{G}=g\sum_{k=1}^{n} x_k x_k^T A^{-1}=g.$$

Таким образом, в случае линейной регрессии $\hat{G}$ представляет собой несмещенную оценку матрицы g.

Для исследования рассеивания оценки $\hat{G}$ обозначим $\hat{G}_1, \ldots, \hat{G}_m$ строки матрицы $\hat{G}$ (m — размерность вектора Y):

$$\hat{G}_p=\sum_{k=1}^{n} Y_{kp} x_k^T A^{-1} \qquad (p=1, \ldots, m),$$

где $Y_{k1}, \ldots, Y_{km}$ — координаты вектора Y_k $(k=1, \ldots, n)$. Найдем ковариационные и взаимные ковариационные матрицы векторов $\hat{G}_p^T$ и $\hat{G}_q^T$:

$$L_{pq}=M\hat{G}_q^{0T}\hat{G}_q^{0}=\sum_{k,l=1}^{n} A^{-1}x_k M Y_{kp}^0 Y_{lq}^0 x_l^T A^{-1}=$$

$$=A^{-1}\sum_{k,l=1}^{n} x_k x_l^T M Y_{kp}^0 Y_{lq}^0 A^{-1}.$$

Так как случайные векторы Y_k и Y_l при $l \neq k$ независимы вследствие независимости опытов, то $MY^0_{kp}Y^0_{lq}=0$ при $l \neq k$. Величина $MY^0_kY^{0T}_k$ представляет собой ковариационную матрицу $K_y(x)$ случайного вектора Y при $x=x_k$. Поэтому

$$MY^0_{kp}Y^0_{kq}=M\left[Y^0_pY^0_q \mid x_k\right]=k_{pq}(x_k).$$

Следовательно,

$$L_{pq}=M\hat{G}^{0T}_p\hat{G}^0_q=A^{-1}\sum_{k=1}^{n}x_kx_k^Tk_{pq}(x_k)A^{-1}. \qquad (31)$$

В частном случае, когда ковариационная матрица K_y вектора Y не зависит от x, формула (31) в силу (28) принимает вид

$$L_{pq}=k_{pq}A^{-1} \qquad (p,\ q=1,\ \ldots,\ m). \blacktriangleleft \qquad (32)$$

Эта формула показывает, что ковариационная матрица случайного вектора $\hat{G}=[\hat{G}_1\ldots\hat{G}_m]^T$ представляет собой прямое (кронекеровское) произведение матриц K_y и A^{-1} [48, 57].

3.3. Оценка ковариационной матрицы наблюдаемой величины. Предположим сначала, что регрессия линейна, а ковариационная матрица K_y величины Y не зависит от x.

▶ В этом случае оценку матрицы K_x можно получить, взяв соответствующие средние значения квадратов и произведений отклонений наблюдаемых значений координат вектора Y от найденной оценки регрессии. В связи с этим рассмотрим статистику

$$S=\sum_{k=1}^{n}(Y_k-\hat{G}x_k)(Y_k-\hat{G}x_k)^T. \qquad (33)$$

Найдем ее математическое ожидание. Заметив, что на основании (27)

$$\sum_{k=1}^{n}(Y_k-\hat{G}x_k)x_k^T=0,$$

перепишем (33) в виде

$$S=\sum_{k=1}^{n}(Y_k-\hat{G}x_k)Y_k^T. \qquad (34)$$

Отсюда находим

$$MS=M\sum_{k=1}^{n}Y_kY_k^T-M\hat{G}\sum_{k=1}^{n}x_kY_k^T. \qquad (35)$$

Величина $MY_kY_k^T$ представляет собой момент второго порядка случайного вектора Y при $x=x_k$. Применяя для его вычисления формулу (3.37), получаем

$$MY_kY_k^T=\Gamma_y(x_k)=m_y(x_k)\,m_y(x_k)^T+K_y=gx_kx_k^Tg^T+K_y.$$

Следовательно,

$$M\sum_{k=1}^{n}Y_kY_k^T=g\sum_{k=1}^{n}x_kx_k^Tg^T+nK_y=gAg^T+nK_y. \qquad (36)$$

Вычислим второе слагаемое в (35). На основании (29)

$$M\hat{G}\sum_{k=1}^{n}x_kY_k^T=M\sum_{k,l=1}^{n}Y_lx_l^TA^{-1}x_kX_k^T.$$

Отсюда, учитывая, что величины $x_l^TA^{-1}x_k$ скалярные, получаем

$$M\hat{G}\sum_{k=1}^{n}x_kY_k^T=\sum_{k,l=1}^{n}x_l^TA^{-1}x_kMY_lY_k^T.$$

Но в силу независимости Y_k и Y_l при $l\neq k$

$$MY_lY_k^T=MY_lMY_k^T=gx_lx_k^Tg^T$$
$$MY_kY_k^T=gx_kx_k^Tg^T+K_y.$$

Следовательно,

$$M\hat{G}\sum_{k=1}^{n}x_kY_k^T=\sum_{k,l=1}^{n}x_k^TA^{-1}x_lgx_kx_l^Tg^T+\sum_{k=1}^{n}x_k^TA^{-1}x_kK_y=$$
$$=\sum_{k,l=1}^{n}gx_kx_k^TA^{-1}x_lx_l^Tg^T+K_y\sum_{k=1}^{n}x_k^TA^{-1}x_k=$$
$$=g\sum_{k=1}^{n}x_kx_k^TA^{-1}\sum_{l=1}^{n}x_lx_l^Tg^T+K_y\,\mathrm{tr}\sum_{k=1}^{n}x_kx_k^TA^{-1}=$$
$$=gAg^T+K_y\,\mathrm{tr}\,AA^{-1}.$$

Но AA^{-1} представляет собой единичную матрицу, порядок которой равен размерности r вектора x. Поэтому $\mathrm{tr}\,AA^{-1}=r$ и, следовательно,

$$M\hat{G}\sum_{k=1}^{n}x_kY_k^T=gAg^T+rK_y.$$

Подставив это выражение и (36) в (35), окончательно получаем

$$MS=(n-r)\,K_y.$$

Отсюда видно, что несмещенная оценка ковариационной матрицы K_y определяется формулой

$$\hat{K}_y = \frac{S}{n-r} = \frac{1}{n-r} \sum_{k=1}^{n} (Y_k - \hat{G}x_k)(Y_k - \hat{G}x_k)^T. \blacktriangleleft \quad (37)$$

Для вычисления величины S часто пользуются формулой

$$S = \sum_{k=1}^{n} Y_k Y_k^T - \hat{G} A \hat{G}^T,$$

вытекающей из (34) и (29). При этом (37) принимает вид

$$\hat{K}_y = \frac{1}{n-r} \left[\sum_{k=1}^{n} Y_k Y_k^T - \hat{G} A \hat{G}^T \right]. \quad (38)$$

Заменив в (32) ковариации k_{pq} их несмещенными оценками из (37), получим несмещенные оценки ковариационных матриц L_{pq} строк матрицы $\hat{G}$.

3.4. Статистические свойства оценок значений регрессии. Изучим теперь статистические свойства оценки регрессии $\hat{Y} = \hat{G}x$ при данном x.

В силу несмещенности оценки $\hat{G}$ величина $\hat{Y} = \hat{G}x$ служит несмещенной оценкой величины gx. Чтобы найти ковариационную матрицу $K_{\hat{y}}$ оценки $\hat{Y}$, заметим, что координаты вектора $\hat{Y}$ выражаются через строки $\hat{G}_1, \ldots, \hat{G}_m$ матрицы $\hat{G}$ формулой $\hat{Y}_p = \hat{G}_p x = x^T \hat{G}_p^T$. Пользуясь этим соотношением и формулой (32), находим элементы матрицы $K_{\hat{y}}$:

$$\varkappa_{pq} = M(x^T \hat{G}_p^{0T})(\hat{G}_q^0 x) = x^T M \hat{G}_p^{0T} \hat{G}_q^0 x = k_{pq} x^T A^{-1} x.$$

Отсюда следует, что

$$K_{\hat{y}} = x^T A^{-1} x K_y. \quad (39)$$

Несмещенная оценка матрицы $K_{\hat{y}}$ получается отсюда заменой матрицы K_y ее несмещенной оценкой из (37).

3.5. Оценивание нелинейной регрессии. Если регрессия нелинейна и ковариационная матрица величины Y зависит от x, то оценку регрессии $m_y(x)$ и ковариационной матрицы $K_y(x)$ можно получить только в том случае, когда при каждом значении x величина Y наблюдается многократно, т. е. $x_1, \ldots, x_n$ распадаются на

группы одинаковых значений *). Предположим, что n_1 из величин $x_1, \ldots, x_n$ равны $x^{(1)}$, n_2 равны $x^{(2)}$, и т. д., n_N равны $x^{(N)}$, $n_1 + \ldots + n_N = n$. Значения величины Y при $x = x^{(p)}$ обозначим через $Y_k^{(p)}$ $(k = 1, \ldots, n_p;\ p = 1, \ldots N)$. Тогда на основании формул (6.34) и (6.37) несмещенные оценки величин $m_y(x)$ и $K_y(x)$ при $x = x^{(1)}, \ldots, x^{(N)}$ определяются формулами

$$\hat{M}_y(x^{(p)}) = \overline{Y}^{(p)} = \frac{1}{n_p} \sum_{k=1}^{n_p} Y_k^{(p)},$$

$$\hat{K}_y(x^{(p)}) = \frac{1}{n_p - 1} \sum_{k=1}^{n_p} (Y_k^{(p)} - \overline{Y}^{(p)})(Y_k^{(p)} - \overline{Y}^{(p)})^T.$$

Статистические характеристики этих оценок и доверительные области в случае нормального распределения Y определяются по формулам § 6.5.

3.6. Случай линейной регрессии и нормального распределения. Рассмотрим подробнее частный случай, когда величина Y имеет нормальное распределение, ее регрессия на x линейна, а ковариационная матрица K_y не зависит от x.

▶ В этом случае совместная плотность наблюдаемых величин $Y_1, \ldots, Y_n$ определяется формулой

$$f(y_1, \ldots, y_n) = \\ = \frac{1}{\sqrt{(2\pi)^{mn} |K_y|^n}} \exp\left\{-\frac{1}{2} \sum_{k=1}^{n} (y_k - gx_k)^T K_y^{-1} (y_k - gx_k)\right\}.$$

Но

$$\sum_{k=1}^{n} (y_k - gx_k)^T K_y^{-1} (y_k - gx_k) = \\ = \operatorname{tr} K_y^{-1} \sum_{k=1}^{n} (y_k - gx_k)(y_k - gx_k)^T.$$

---

*) Оценки величин $m_y(x)$ и $K_y(x)$ можно найти методом наименьших квадратов при однократном наблюдении Y при каждом из значений $x_1, \ldots, x_n$ величины x, но только в классе достаточно простых функций, число неизвестных параметров в которых значительно меньше числа опытов n.

Сумма в этой формуле выражается через величины $\hat{g}$ и s. Действительно,

$$\sum_{k=1}^{n}(y_k-gx_k)(y_k-gx_k)^T=$$
$$=\sum_{k=1}^{n}[y_k-\hat{g}x_k+(\hat{g}-g)x_k][y_k-\hat{g}x_k+(\hat{g}-g)x_k]^T=$$
$$=\sum_{k=1}^{n}(\hat{g}-g)x_kx_k^T(\hat{g}-g)^T+\sum_{k=1}^{n}(\hat{g}-g)x_k(y_k^T-x_k^T\hat{g}^T)+$$
$$+\sum_{k=1}^{n}(y_k-\hat{g}x_k)x_k^T(\hat{g}-g)^T+\sum_{k=1}^{n}(y_k-\hat{g}x_k)(y_k-\hat{g}x_k)^T.$$

Отсюда, принимая во внимание, что на основании (27) и (28)

$$\sum_{k=1}^{n}(y_k-\hat{g}x_k)x_k^T=0,\quad \sum_{k=1}^{n}x_k(y_k-gx_k)^T=0,\quad \sum_{k=1}^{n}x_kx_k^T=A,$$

получаем

$$\sum_{k=1}^{n}(y_k-gx_k)(y_k-gx_k)^T=$$
$$=(\hat{g}-g)A(\hat{g}-g)^T+\sum_{k=1}^{n}(y_k-\hat{g}x_k)(y_k-\hat{g}x_k)^T=$$
$$=(\hat{g}-g)A(\hat{g}-g)^T+s.$$

Следовательно,

$$\sum_{k=1}^{n}(y_k-gx_k)^TK_y^{-1}(y_k-gx_k)=$$
$$=\operatorname{tr}K_y^{-1}(\hat{g}-g)A(\hat{g}-g)^T+\operatorname{tr}K_y^{-1}s$$

и

$$f(y_1,\ldots,y_n)=\frac{1}{\sqrt{(2\pi)^{mn}|K_y|^n}}\times$$
$$\times\exp\left\{-\frac{1}{2}\operatorname{tr}K_y^{-1}(\hat{g}-g)A(\hat{g}-g)^T-\frac{1}{2}\operatorname{tr}K_y^{-1}s\right\}.\quad (40)$$

Отсюда следует, что случайные величины $\hat{G}$ и S независимы и образуют достаточную статистику, причем $\hat{G}$

распределена нормально, а S имеет распределение Уишарта $w_{n-r+1,\,m}(s)$. Уменьшение числа степеней свободы S от $n-1$ до $n-r$ по сравнению с S в § 6.5 объясняется тем, что векторы $Y_k-\hat{G}x_k$ $(k=1, \ldots, n)$ связаны r линейными зависимостями

$$\sum_{k=1}^{n}(Y_k-\hat{G}x_k)\,x_{kp}=0 \qquad (p=1, \ldots, r),$$

вытекающими из (27) *). Из (40), так же как в примере 7.6, вытекает, что оценками максимума правдоподобия параметров g и K_y служат величины $\hat{G}$ и $S/n=\hat{K}_y(n-r)/n$. ◀

3.7. Выбор значений независимой переменной. Если значения $x_1, \ldots, x_n$ величины x, при которых наблюдается величина Y, можно выбирать произвольно, то вычисление оценки $\hat{G}$ можно упростить, выбрав $x_1, \ldots, x_n$ так, чтобы матрица A была диагональной. Тогда уравнение (27) распадается на mr независимых уравнений, определяющих элементы матрицы $\hat{G}$ по отдельности.

Чтобы найти векторы $x_1, \ldots, x_n$ с диагональной матрицей A, возьмем произвольные векторы $x'_1, \ldots, x'_n$ и обозначим через u'_p n-мерный вектор, образованный p-ми координатами векторов $x'_1, \ldots, x'_n$:

$$u'_p=[x'_{1p} \ldots x'_{np}]^T \qquad (p=1, \ldots, r).$$

Тогда можно будет рассматривать элементы матрицы A как скалярные произведения соответствующих векторов u'_p и u'_q. Применив к векторам $u'_1, \ldots, u'_r$ обычный прием ортогонализации, заменим их ортогональными векторами $u_1, \ldots, u_r$. Обозначив через x_k вектор, образованный k-ми координатами векторов $u_1, \ldots, u_r$ $(k=1, \ldots, n)$, получим векторы $x_1, \ldots, x_n$ с диагональной матрицей A. Само собой разумеется, что исходные векторы $x'_1, \ldots, x'_n$ следует выбирать так, чтобы векторы $x_1, \ldots, x_n$ были распределены приблизительно равномерно в интересующей нас области значений x.

3.8. Доверительные области для регрессии. Осталось найти доверительные области для g и K_y.

---

*) В частном случае § 6.5, когда $r=1$, $x_1=1$, векторы $Y_k-\hat{G}x_k=Y_k-\overline{Y}$, $\overline{Y}=(1/n)\sum Y_k$, связаны одним соотношением $\sum(Y_k-\overline{Y})=0$.

▶ Для определения доверительной области для регрессии $y = gx$ при данном значении x воспользуемся третьим методом п. 6.2.4. Так же как в примерах 5.15 и 5.24, убеждаемся в том, что случайная величина

$$F = \frac{T^2}{n-r} \frac{n-r-m+1}{m} = \frac{n-r-m+1}{mx^T A^{-1} x} (\hat{Y} - y)^T S^{-1} (\hat{Y} - y)$$

имеет F-распределение $f_{m,\, n-r-m+1}(f)$, не зависящее от неизвестного параметра g. Поэтому величину F можно принять за функцию $\varphi(\hat{Y}, S, g)$ п. 6.2.4. Тогда доверительная область для значения $y = gx$ регрессии при данном x определится неравенством

$$(\hat{Y} - y)^T S^{-1} (\hat{Y} - y) < f_\alpha \frac{mx^T A^{-1} x}{n-r-m+1}, \qquad (41)$$

где f_α — верхняя $100(1-\alpha)$-процентная точка F-распределения $f_{m,\, n-r-m+1}(f)$ (значение f, при котором функция распределения равна α).

В частном случае скалярной величины Y распределение $f_{1,\, n-r}(f)$ величины F дает для величины $T = +\sqrt{F}$ плотность $2s_{n-r}(t)$. Доверительная область (41) для $y = gx$ в этом случае представляет собой интервал

$$\hat{Y} - t_\alpha \sqrt{\frac{S}{n-r} x^T A^{-1} x} < y < \hat{Y} + t_\alpha \sqrt{\frac{S}{n-r} x^T A^{-1} x},$$

где t_α определяется уравнением $S_{n-r}(t_\alpha) = P(|T| < t_\alpha) = \alpha$ (таблица 3). ◀

Доверительная область для матрицы g в случае векторной величины Y определяется сложно, и удобных для практического применения результатов при этом получить не удается. Путь решения этой задачи будет показан в § 4.

▶ В частном случае скалярной величины Y распределение Уишарта величины S сводится к χ^2-распределению с $n-r$ степенями свободы величины S/D_y. А так как независимая от S величина $(\hat{G} - g) A (\hat{G} - g)^T / D_y$ на основании результатов примера 5.24 имеет χ^2-распределение с r степенями свободы, то величина

$$F = \frac{(\hat{G} - g) A (\hat{G} - g)^T / r}{S / n-r}$$

имеет F-распределение $f_{r,\, n-r}(f)$. Следовательно, доверительная область для матрицы-строки g определяется

неравенством

$$(\hat{G}-g)\, A\, (\hat{G}-g)^T < f_\alpha \frac{rS}{n-r}, \qquad (42)$$

где f_α — верхняя $100(1-\alpha)$-процентная точка F-распределения $f_{r,\,n-r}(f)$. ◀

В случае векторной величины Y неравенство (42) определяет доверительные области для строк матрицы g по отдельности.

3.9. Оценивание линейной регрессии со сдвигом. В задачах практики часто приходится находить линейную оценку регрессии со сдвигом.

▶ В случае линейной оценки регрессии со сдвигом первая координата вектора x равна 1 в соответствии с замечанием п. 2.2.5. Заменив в этом случае g, $\hat{g}$ и x соответственно блочными матрицами $[a\ g]$, $[\hat{a}\ \hat{g}]$ и $[1\ x^T]^T$, напишем уравнение (27) в виде

$$[\hat{a}\ \hat{g}] \begin{bmatrix} n & \sum_{k=1}^{n} x_k^T \\ \sum_{k=1}^{n} x_k & \sum_{k=1}^{n} x_k x_k^T \end{bmatrix} = \left[\sum_{k=1}^{n} y_k \quad \sum_{k=1}^{n} y_k x_k^T \right].$$

Это уравнение распадается на два:

$$n\hat{a} + \hat{g} \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k,$$

$$\hat{a} \sum_{k=1}^{n} x_k^T + \hat{g} \sum_{k=1}^{n} x_k x_k^T = \sum_{k=1}^{n} y_k x_k^T.$$

Первое уравнение дает

$$\hat{a} = \frac{1}{n} \sum_{k=1}^{n} y_k - \frac{\hat{g}}{n} \sum_{k=1}^{n} x_k = \bar{y} - \hat{g}\bar{x}.$$

Подставив это выражение во второе уравнение, приведем его к виду

$$\hat{g} \left(\sum_{k=1}^{n} x_k x_k^T - n\bar{x}\bar{x}^T \right) = \sum_{k=1}^{n} y_k x_k^T - n\bar{y}\bar{x}^T,$$

или

$$\hat{g} \sum_{k=1}^{n} (x_k - \bar{x})(x_k - \bar{x})^T = \sum_{k=1}^{n} (y_k - \bar{y})(x_k - \bar{x})^T. \qquad (43)$$

Таким образом, в этом случае можно представить оптимальную линейную оценку регрессии в виде

$$\hat{y}(x)=\overline{y}+\hat{g}(x-\overline{x}),$$

где $\hat{g}$ определяется уравнением (43). Это дает возможность уменьшить на 1 размерность матрицы коэффициентов и число столбцов матрицы в правой части по сравнению с (27).

Доверительная область (41) для значения регрессии $a+gx=y$ при данном x преобразуется в этом случае к виду

$$(\overline{Y}+\hat{G}(x-\overline{x})-y)^T S^{-1}(\overline{Y}+\hat{G}(x-\overline{x})-y)<$$
$$< f_\alpha \frac{m}{n-r-m+1}\left[\frac{1}{n}+(x-\overline{x})^T A'^{-1}(x-\overline{x})\right],$$

где

$$A'=\sum_{k=1}^{n}(x_k-\overline{x})(x_k-\overline{x})^T.$$

Доверительная область (42) для матрицы-строки $[a\ g]$ в случае скалярной величины Y имеет вид

$$n(\overline{Y}-a-g\overline{x})^2+(\hat{G}-g)A'(\hat{G}-g)^T<f_\alpha\frac{rS}{n-r}.$$

При этом следует помнить, что векторы $x_1, \ldots, x_n$ и $\overline{x}$ в этом случае имеют размерность $r-1$. ◀

Чтобы получить решение уравнения (43) в удобной для вычислений форме, обозначим через $\hat{\sigma}_x$ и $\hat{\sigma}_y$ диагональные матрицы, элементами которых служат соответственно выборочные средние квадратические отклонения координат векторов Y и x, через $\hat{R}_x$ и $\hat{R}_{yx}$ выборочные корреляционную матрицу вектора x и взаимную корреляционную матрицу векторов Y и x *). Тогда можем написать

$$\begin{gathered}A'=\sum_{k=1}^{n}(x_k-\overline{x})(x_k-\overline{x})^T=(n-1)\hat{\sigma}_x\hat{R}_x\hat{\sigma}_x,\\ \sum_{k=1}^{n}(y_k-\overline{y})(x_k-\overline{x})^T=(n-1)\hat{\sigma}_y\hat{R}_{yx}\hat{\sigma}_x,\end{gathered}\qquad(44)$$

и уравнение (43) примет вид

$$\hat{g}\hat{\sigma}_x\hat{R}_x=\hat{\sigma}_y\hat{R}_{yx}.$$

*) Выборочные статистические характеристики, очевидно, могут быть вычислены и в том случае, когда наблюдаемые величины не случайны.

Решая это уравнение, получим

$$\hat{g}=\hat{\sigma}_y\hat{R}_{yx}\hat{R}_x^{-1}\hat{\sigma}_x^{-1}. \tag{45}$$

Очевидно, что строки матрицы $\hat{g}$ определяются независимо одна от другой вследствие того, что уравнение (43) (как и (27)) распадается на независимые уравнения такого же вида, каждое из которых определяет соответствующую строку матрицы g. Вследствие этого регрессию можно оценивать для каждой координаты вектора Y отдельно. Этим обычно пользуются при вычислениях.

В случае скалярной величины Y (или одной координаты Y интересующего нас случайного вектора) для вычисления реализации s статистики S удобно применить следующий прием. Так же как при переходе от формулы (37) к (38), получаем

$$s=\sum_{k=1}^{n}[y_k-\overline{y}-\hat{g}(x_k-\overline{x})]^2=\sum_{k=1}^{n}(y_k-\overline{y})^2-\hat{g}A'\hat{g}^T.$$

Подставив сюда выражения A' и $\hat{g}$ из (44) и (45) и учитывая, что $R_{yx}^T=R_{xy}$, находим

$$s=\sum_{k=1}^{n}(y_k-\overline{y})^2-(n-1)\hat{\sigma}_y^2R_{yx}R_x^{-1}R_{xy},$$

или, полагая

$$s'=\sum_{k=1}^{n}(y_k-\overline{y})^2$$

и принимая во внимание, что $(n-1)\hat{\sigma}_y^2=s'$,

$$s=s'(1-R_{yx}R_x^{-1}R_{xy}). \tag{46}$$

Формула (37) дает несмещенную оценку дисперсии D_y величины Y:

$$\hat{d}_y=\frac{s}{n-r}. \tag{47}$$

Наконец, учитывая, что в данном случае

$$A^{-1}=\begin{bmatrix}\frac{1}{n}+\overline{x}^TA'^{-1}\overline{x} & -\overline{x}^TA'^{-1}\\ -A'^{-1}\overline{x} & A'^{-1}\end{bmatrix},$$

получаем из (32) несмещенную оценку ковариационной матрицы оценки параметра $[a\ g]$:

$$\hat{d}_a=\hat{d}_y\left(\frac{1}{n}+\overline{x}^TA'^{-1}\overline{x}\right),\quad \hat{k}_{ga}=-\hat{d}_yA'^{-1}\overline{x},\quad \hat{k}_g=\hat{d}_yA'^{-1}.$$

Подставив сюда выражение A'^{-1} из (44), получаем

$$\begin{gathered}\hat{d}_a=\frac{\hat{d}_y}{n}+\overline{x}^T\hat{k}_g\overline{x},\quad \hat{k}_{ga}=-\hat{k}_g\overline{x},\\ \hat{k}_g=\hat{d}_y\frac{\hat{\sigma}_x^{-1}\hat{R}_x^{-1}\hat{\sigma}_x^{-1}}{n-1}.\end{gathered} \tag{48}$$

Формулы (45) — (48) лежат в основе стандартной программы линейного регрессионного анализа MULTR [94] *).

Пример 4 **). Построить оптимальную линейную регрессионную модель по результатам 30 наблюдений величины Y и пятимерного вектора x, представленным в таблице:

k	x_k					y_k	k	x_k					y_k
1	29	289	216	85	14	1	16	45	310	262	94	20	2
2	30	391	244	92	16	2	17	45	151	339	96	35	3
3	30	424	246	90	18	2	18	45	370	357	88	15	4
4	30	313	239	91	10	0	19	45	379	198	147	64	4
5	35	243	275	95	30	2	20	45	463	206	105	31	3
6	35	365	219	95	21	2	21	45	316	245	132	60	4
7	43	396	267	100	39	3	22	45	280	225	108	36	4
8	43	356	274	79	19	2	23	44	395	215	101	27	1
9	44	346	255	126	56	3	24	49	139	220	136	59	0
10	44	156	258	95	28	0	25	49	245	205	113	37	4
11	44	278	249	110	42	4	26	49	373	215	88	25	1
12	44	349	252	88	21	1	27	51	224	215	118	54	3
13	44	141	236	129	56	1	28	51	677	210	116	33	4
14	44	245	236	97	24	1	29	51	424	210	140	59	4
15	45	297	256	111	45	3	30	51	150	210	105	30	0

В результате расчета на ЭВМ по программе REGRE получена зависимость

$$\hat{y}(x) = 2{,}267 + 0{,}01\,[1{,}242\ \ 0{,}739\ \ 1{,}504\ \ 0{,}151\ \ 4{,}919] \times$$
$$\times\,(x - [43{,}13\ \ 316{,}17\ \ 241{,}80\ \ 105{,}67\ \ 34{,}13]^T$$

и оценки средних квадратических отклонений величины Y и оценок коэффициентов регрессии

$$\hat{\sigma}_y = 1{,}05162, \quad \hat{\sigma}_{g_1} = 0{,}03635, \quad \hat{\sigma}_{g_2} = 0{,}00186,$$
$$\hat{\sigma}_{g_3} = 0{,}00635, \quad \hat{\sigma}_{g_4} = 0{,}03679, \quad \hat{\sigma}_{g_5} = 0{,}04141.$$

Анализ этих результатов показывает, что средние квадратические отклонения оценок первого и четвертого коэффициентов регрессии значительно превышают оценки этих коэффициентов. Это приводит к гипотезе, что зависимость регрессии от первой и четвертой координат вектора x несущественна. В примере 8 мы увидим, что эта гипотеза может быть принята.

---

*) Для построения линейных регрессионных моделей в [94] служит общая программа REGRE, использующая программу CORRE для вычисления выборочных средних, дисперсий, ковариаций, коэффициентов корреляции и сумм квадратов и произведений, программу MINV для обращения матрицы R_x и программу MULTR для непосредственного построения регрессионной модели.

**) Примеры 4 и 5 заимствованы из [94].

Принимая во внимание, что в данном случае $n=30$, $r=6$, $s=26{,}54161$, и определив по таблице 6 верхнюю 5%-ную точку F-распределения $f_{6,24}(f)$, $f_{0,95}=2{,}51$, находим доверительную область для матрицы $[a\ g]$, соответствующую коэффициенту доверия 0,95:

$$(2{,}267-a-g\bar{x})^2+(\hat{g}-g)\,A''\,(\hat{g}-g)^T<0{,}555,$$

где

$$\bar{x}=[43{,}13\ \ 316{,}17\ \ 241{,}80\ \ 105{,}67\ \ 34{,}13]^T,$$

$$\hat{g}=0{,}01\,[1{,}242\ \ 0{,}739\ \ 1{,}504\ \ 0{,}151\ \ 4{,}919]^T,$$

$$A''=\frac{A'}{n}=\frac{n-1}{n}\hat{\sigma}_x\hat{R}_x\hat{\sigma}_x=\frac{29}{30}\hat{\sigma}_x\hat{R}_x\hat{\sigma}_x,$$

$\hat{\sigma}_x$ — диагональная матрица с элементами 6,52176, 114,42990, 36,43074, 17,85640, 15,97571, а

$$\hat{R}_x=\begin{bmatrix}1{,}00000 & -0{,}06721 & -0{,}13689 & 0{,}49755 & 0{,}55849\\ -0{,}06721 & 1{,}00000 & -0{,}17857 & -0{,}05227 & -0{,}18381\\ -0{,}13689 & -0{,}17857 & 1{,}00000 & -0{,}40874 & -0{,}26319\\ 0{,}49755 & -0{,}05227 & -0{,}40874 & 1{,}00000 & 0{,}93552\\ 0{,}55849 & -0{,}18381 & -0{,}26319 & 0{,}93552 & 1{,}00000\end{bmatrix}.$$

Доверительная область для значения линейной регрессии y при данном x определяется неравенством

$$y-2{,}267-\hat{g}(x-\bar{x})\,|<$$
$$<0{,}396\sqrt{1+1{,}0345\,(x-\bar{x})^T\hat{\sigma}_x^{-1}\hat{R}_x^{-1}\hat{\sigma}_x^{-1}(x-\bar{x})}$$

при тех же $\bar{x}$, $\hat{g}$, $\hat{\sigma}_x$ и $\hat{R}_x$ после определения по таблице 3 значения $t_\alpha=t_{0,95}=2{,}064$ по данным $k=n-r=24$ и $\alpha=0{,}95$.

Пример 5. Найти полиномиальную оценку регрессии 4-й степени скалярной величины Y на скалярную величину x по наблюденным значениям x и Y в 15 опытах, приведенным в таблице:

x_k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y_k	10	16	20	23	25	26	30	36	48	62	78	94	107	118	127

В результате расчета на ЭВМ по программе POLGR *) получена зависимость

$$\hat{y}(x)=-5{,}642+19{,}449x-5{,}1147x^2+0{,}56508x^3-0{,}018113x^4.$$

*) В [94] для построения полиномиальной регрессионной модели служит программа POLGR, использующая программы GDATA для вычисления степеней независимой переменной — координат вектора z, их выборочных средних, сумм квадратов, сумм произведений и коэффициентов корреляции, MINV для обращения корреляционной матрицы и MULTR для непосредственного построения модели.

3.10. Оценивание регрессии одной случайной величины на другую. Перейдем теперь к случаю, когда x представляет собой значение некоторой случайной величины X. Задача нахождения моментов оценок и доверительных областей для a, g и $y=a+gx$ в этом случае значительно труднее. При нормальном совместном распределении величин X и Y сравнительно легко находятся лишь доверительные интервалы для элементов матрицы g по отдельности. Для решения этой задачи достаточно рассмотреть случай скалярной величины Y *).

▶ Выделим одну первую координату вектора X, которую мы обозначим $X_{\cdot 1}$. Точка перед индексом поставлена для того, чтобы отличить первую координату вектора X от его значения в первом опыте X_1. Представим вектор X в виде блочной матрицы-столбца $X = [X_{\cdot 1} \;\; X'^T]^T$ **). Соответственно представим матрицу-строку g в виде $g=[g_1 \;\; g']$ и значения вектора X в соответствующих опытах в виде $X_k=[X_{k1} \; X_k'^T]^T$ $(k=1, \ldots, n)$. Пусть g — матрица-строка коэффициентов регрессии Y на $X=[X_{\cdot 1} \;\; X'^T]^T$, h_1 — матрица-строка коэффициентов регрессии Y на X', h_2 — матрица-строка коэффициентов регрессии $X_{\cdot 1}$ на X', $\hat{G}$, $\hat{H}_1$, $\hat{H}_2$ — оптимальные оценки матриц g, h_1, h_2. Положим

$$
\begin{aligned}
S_{11} &= \sum_{k=1}^{n} [Y_k - \bar{Y} - \hat{H}_1 (X_k' - \bar{X}')]^2, \quad \bar{X}' = \frac{1}{n} \sum_{k=1}^{n} X_k', \\
S_{22} &= \sum_{k=1}^{n} [X_{k1} - \bar{X}_{\cdot 1} - \hat{H}_2 (X_k' - \bar{X}')]^2, \quad \bar{X}_{\cdot 1} = \frac{1}{n} \sum_{k=1}^{n} X_{k1}, \qquad (49) \\
S_{12} &= \sum_{k=1}^{n} [Y_k - \bar{Y} - \hat{H}_1 (X_k' - \bar{X}')] [X_{k_1} - \bar{X}_{\cdot 1} - \hat{H}_2 (X_k' - \bar{X}')].
\end{aligned}
$$

Бартлетт доказал, что статистика

$$
T = \sqrt{\frac{n-r-1}{S_{11} S_{22} - S_{12}^2}} \, S_{11} (\hat{G}_1 - g_1) \qquad (50)
$$

---

*) Впрочем, все полученные выше результаты в этом случае справедливы для условных характеристик при данных $x_1, \ldots, x_n$.

**) Чтобы избежать вырожденного распределения, в этом случае регрессия Y на X всегда представляется в виде $\hat{y}(x) = \bar{y} + g(x - \bar{x})$.

имеет T-распределение $s_{n-r-1}(t)$ [2]. Это дает возможность находить доверительные интервалы для коэффициента регрессии g_1. Совершенно так же определяются доверительные интервалы для $g_2, \ldots, g_r$. ◀

Пример 6. Найти регрессию случайной величины Y на X в условиях примера 6.7.

По формуле (45), имея в виду, что в данном случае $n=20$, $r=1$, $\hat{R}_x=1$, $\hat{R}_{yx}=\hat{r}_{xy}$, находим $\hat{g}=0{,}57$. Так как в данном случае $s_{11}=(n-1)\hat{\sigma}_x^2$, $s_{22}=(n-1)\hat{\sigma}_y^2$, $s_{12}=(n-1)\hat{k}_{xy}=(n-1)\hat{\sigma}_x\hat{\sigma}_y\hat{r}_{xy}$, то доверительный интервал для g определяется неравенством

$$\frac{\hat{\sigma}_x}{\hat{\sigma}_y}\sqrt{\frac{n-2}{1-\hat{r}_{xy}^2}}\,|\hat{g}-g| < t_\alpha,$$

где t_α определяется уравнением $S_{n-2}(t_\alpha)=\alpha$. Полагая $\alpha=0{,}95$ и учитывая, что $k=n-2=18$, находим по таблице 3 $t_\alpha=t_{0,95}=2{,}103$. После этого определяем доверительный интервал $0{,}27 < g < 0{,}87$.

§ 4. Проверка гипотез о регрессии

4.1. Критерий для проверки гипотез о равенстве нулю коэффициентов регрессии. При построении регрессионных моделей обычно возникает вопрос, нельзя ли пренебречь зависимостью Y от некоторых координат вектора x. Иными словами, возникает вопрос о проверке гипотезы, что некоторые элементы матрицы коэффициентов регрессии g равны нулю.

Пусть x' — вектор, образованный первыми p координатами вектора x, x'' — вектор, образованный остальными $r-p$ координатами. Представив матрицу-столбец x в виде блочной матрицы $x=[x'^T\ x''^T]^T$ и матрицу g в соответствующем блочном виде $g=[g'\ g'']$, будем иметь $gx=g'x'+g''x''$. Если регрессия не зависит от x'', то $g''=0$. Задача заключается в том, чтобы определить, согласуется ли гипотеза $g''=0$ с полученными экспериментальными данными.

Чтобы найти подходящий критерий для проверки гипотезы $g''=0$, рассмотрим сначала случай скалярной величины Y. Предположим, что найдены оценки регрессии при $g''=0$ и без этого допущения и в обоих случаях вычислены суммы квадратов отклонений экспериментальных точек от регрессии, которые мы обозначим соответственно S' и S. Ясно, что если гипотеза $g''=0$ верна, то S и S' должны быть близкими, хотя S' всегда несколько больше, чем S. Если же гипотеза $g''=0$ неверна, то S'

может быть значительно больше S. Таким образом, в качестве меры отклонения экспериментальных данных от гипотезы $g''=0$ можно принять отношение $U=S/S'$. Если найденное по экспериментальным данным значение U близко к 1, то можно считать, что гипотеза $g''=0$ не противоречит экспериментальным данным. При малом значении U гипотеза $g''=0$ плохо согласуется с экспериментальными данными и должна быть отвергнута.

В случае векторной величины Y отклонения экспериментальных точек от регрессии характеризуются матрицей, элементами которой служат суммы квадратов и произведений отклонений координат экспериментальных точек от регрессии, т. е. матрицей S § 3. При этом величину рассеивания экспериментальных точек относительно регрессии можно характеризовать определителем матрицы S (который пропорционален объему эллипсоида рассеивания). Поэтому за критерий для проверки гипотезы $g''=0$ в случае векторной величины Y обычно принимают отношение определителя матрицы S, определяемой формулой (33), к определителю такой же матрицы S', вычисленной при гипотезе $g''=0$, $U=|S|/|S'|$.

4.2. Вспомогательное предложение. Вычислим разность $S'-S$. Пусть $\hat{G}'$ — оценка матрицы g' при гипотезе $g''=0$, $\hat{G}^{(1)}$ и $\hat{G}^{(2)}$ — оценки матриц g' и g'', полученные без допущения $g''=0$, $\hat{G}=[\hat{G}^{(1)}\hat{G}^{(2)}]$.

▶ Представив матрицы A и B в соответствующем блочном виде:

$$A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad B=[B' \ B''],$$

можем переписать уравнение (27) для оценки $\hat{G}$ в виде двух уравнений:

$$\hat{G}^{(1)}A_{11}+\hat{G}^{(2)}A_{21}=B', \qquad \hat{G}^{(1)}A_{12}+\hat{G}^{(2)}A_{22}=B''. \qquad (51)$$

Уравнение (27) для оценки $\hat{G}'$ имеет вид

$$\hat{G}'A_{11}=B'. \qquad (52)$$

Выразим матрицы S и S' формулой (34):

$$S=\sum_{k=1}^{n}(Y_k-\hat{G}x_k)Y_k^T=\sum_{k=1}^{n}(Y_k-\hat{G}^{(1)}x'_k-\hat{G}^{(2)}x''_k)Y_k^T,$$

$$S'=\sum_{k=1}^{n}(Y_k-\hat{G}'x'_k)Y_k^T.$$

Вычитая первую формулу из второй, получим

$$S'-S=\sum_{k=1}^{n}\hat{G}^{(2)}x_k''Y^T+\sum_{k=1}^{n}(\hat{G}^{(1)}-\hat{G}')\,x_k'Y_k^T=$$
$$=\hat{G}^{(2)}B''^T+(\hat{G}^{(1)}-\hat{G}')\,B'^T. \quad (53)$$

Чтобы вычислить разность $\hat{G}^{(1)}-\hat{G}'$, вычтем уравнение (52) из первого уравнения (51):

$$(\hat{G}^{(1)}-\hat{G}')\,A_{11}+\hat{G}^{(2)}A_{21}=0.$$

Отсюда находим

$$\hat{G}^{(1)}-\hat{G}'=-\hat{G}^{(2)}A_{21}A_{11}^{-1}.$$

Подставив это выражение в (53), получаем

$$S'-S=\hat{G}^{(2)}(B''^T-A_{21}A_{11}^{-1}B'^T). \quad (54)$$

Для дальнейшего преобразования этого выражения исключим из (51) матрицу $\hat{G}^{(1)}$. С этой целью умножим первое уравнение (51) на $A_{11}^{-1}A_{12}$ и вычтем из второго:

$$\hat{G}^{(2)}(A_{22}-A_{21}A_{11}^{-1}A_{12})=\hat{B}''-B'A_{11}^{-1}A_{12}.$$

Отсюда, имея в виду, что $A_{21}=A_{12}^T$, $A_{11}^T=A_{11}$, $A_{22}^T=A_{22}$ вследствие симметрии матрицы A, находим

$$B''^T-A_{21}A_{11}^{-1}B'^T=(A_{22}-A_{21}A_{11}^{-1}A_{12})\,\hat{G}^{(2)\,T}.$$

Подставив это выражение в (54), окончательно получаем

$$S'-S=\hat{G}^{(2)}(A_{22}-A_{21}A_{11}^{-1}A_{12})\,\hat{G}^{(2)\,T}. \blacktriangleleft \quad (55)$$

По доказанному в п. 3.6 при данных значениях $x_1, \ldots, x_n$ и при нормальном распределении величины Y статистики $\hat{G}$ и S независимы. Формула (55) показывает, что матрица S' в этом случае представляет собой сумму двух независимых положительно определенных матриц S и $S'-S=\hat{G}^{(2)}(A_{22}-A_{21}A_{11}^{-1}A_{12})\,\hat{G}^{(2)\,T}$.

4.3. Проверка гипотез в случае скалярной наблюдаемой величины. Рассмотрим подробнее случай скалярной величины Y. В этом случае величина S/D_y имеет χ^2-распределение с $n-r$ степенями свободы и, соответственно, S'/D_y имеет χ^2-распределение с $n-p$ степенями свободы. Следовательно, независимая от S величина $(S'-S)/D_y$ имеет χ^2-распределение с $n-p-(n-r)=$ $=(r-p)$ степенями свободы. Отсюда следует, что вели-

чина $U=S/S'$ имеет β-распределение с параметрами $(n-r)/2$ и $(r-p)/2$, а величина

$$F=\frac{(S'-S)/(r-p)}{S/(n-r)}$$

имеет F-распределение с $r-p$ и $n-r$ степенями свободы $f_{r-p,\,n-r}(f)$ (примеры 5.24 и 5.26). Поэтому для проверки гипотезы $g''=0$ можно пользоваться таблицами F-распределения. По этим таблицам определяется верхняя $100(1-\alpha)$-процентная точка f_α F-распределения $f_{r-p,\,n-r}(f)$, т. е. такое значение f_α, при котором $P(F<f_\alpha)=\alpha$. Если полученное экспериментально значение f величины F значительно меньше f_α, то гипотезу $g''=0$ можно считать согласующейся с экспериментальными данными. Для подкрепления решения, что $g''=0$, в этом случае можно, пользуясь F-распределением или β-распределением, вычислить вероятность того, что F примет значение большее, чем экспериментально полученное значение f (соответственно вероятность того, что U примет значение меньшее, чем экспериментально полученное $u=1/[1+f(r-p)/(n-r)]$). Если эта вероятность велика, то это служит дополнительным доводом в пользу гипотезы $g''=0$. При $f>f_\alpha$ гипотезу $g''=0$ следует отвергнуть.

Пример 7. В условиях примера 4 проверить гипотезу $g=0$.

В этом случае $n=30$, $r=6$, $p=1$ и верхняя 5%-ная точка F-распределения $f_{5,24}(f)$ есть $f_{0,95}=2{,}62$. Экспериментально полученное значение f определяется формулой

$$f=\frac{(s'-s)/5}{s/24}, \quad s'=\sum_{k=1}^{30}(y_k-\bar{y})^2.$$

Это значение вычисляется подпрограммой MULTR. В данном случае оно равно 5,6608, т. е. больше чем в два раза превышает критическое значение $f_{0,95}=2{,}62$. Поэтому гипотеза $g=0$ должна быть отвергнута.

Пример 8. В условиях примера 4 проверить гипотезу, что регрессия не зависит от первой и четвертой координат вектора x.

В этом случае $n=30$, $r=6$, $p=4$ и в результате расчета на ЭВМ по программе MULTR получено $s=26{,}54161$, $s'=26{,}67073$. Следовательно,

$$f=\frac{(s'-s)/(r-p)}{s/(n-r)}=\frac{0{,}12912/2}{26{,}54161/24}=0{,}05837.$$

Это значение почти в 60 раз меньше критического значения f_α F-распределения $f_{2,24}(f)$, соответствующего $\alpha=0{,}95$, $f_{0,95}=3{,}40$. Поэтому гипотезу $g_1=g_4=0$ можно принять. Оптимальная регрессионная модель при этой гипотезе определяется формулой

$$\hat{y}(x)=2{,}267+0{,}01\,[0{,}744\ \ 1{,}497\ \ 5{,}363]\,(x'-[316{,}17\ \ 241{,}80\ \ 34{,}13]^T).$$

При этом получены следующие оценки средних квадратических отклонений величины Y и оценок коэффициентов регрессии g_2, g_3 и g_5:

$$\hat{\sigma}_y = 1{,}01282, \quad \hat{\sigma}_{g_2} = 0{,}00172, \quad \hat{\sigma}_{g_3} = 0{,}00551, \quad \hat{\sigma}_{g_5} = 0{,}01258.$$

Для нахождения доверительной области для a, g_2, g_3 и g_5, $[g_2 \; g_3 \; g_5] = g'$, определяем верхнюю 5%-ную точку F-распределения $f_{4,26}(f)$, $f_{0,95} = 2{,}74$. После этого доверительная область для a и g' определится неравенством

$$(2{,}267 - \hat{g}'\bar{x}' - a)^2 + (\hat{g}' - g')\, A''\, (\hat{g}' - g')^T < 0{,}375,$$

где

$$\bar{x}' = [316{,}17 \; 241{,}80 \; 34{,}13]^T, \quad \hat{g}' = 0{,}01\,[0{,}744 \;\; 1{,}497 \;\; 5{,}363],$$

а матрица A'' получается из такой же матрицы примера 4 вычеркиванием первой и четвертой строк и первого и четвертого столбцов.

Обратим внимание на то, что в этой упрощенной модели средние квадратические отклонения оценок $\hat{g}_2$, $\hat{g}_3$ и $\hat{g}_5$ меньше, чем в полной модели примера 4, причем $\hat{\sigma}_{g_5}$ примерно в 3 раза меньше, чем в полной модели.

4.4. Проверка гипотез в случае векторной наблюдаемой величины. Перейдем к случаю m-мерной векторной величины Y. Рассмотрим сначала случай $p = r - 1$, т. е. случай проверки гипотезы о независимости Y от одной r-й координаты вектора x. В этом случае величина $U = |S| / |S'|$ имеет β-распределение с параметрами $(n - p - m)/2$, $m/2$ и, следовательно, величина

$$F = \frac{(1 - U)/m}{U/(n - p - m)}$$

имеет F-распределение с m и $n - p - m$ степенями свободы $f_{m,\, n-p-m}(f)$. Поэтому гипотеза $g'' = 0$ проверяется совершенно так же, как в случае скалярной Y.

В случае, когда $r - m < p < r - 1$, величина U может быть представлена в виде произведения $r - p$ независимых случайных величин $X_1, \ldots, X_{r-p}$:

$$U = \prod_{l=1}^{r-p} X_l,$$

где X_l имеет β-распределение с параметрами $(n - p - m - l + 1)/2$, $m/2$. Это доказывается путем вычисления моментов величины U совершенно так же, как в примере 5.45 [1].

При $p \leqslant r - m$, что, конечно, возможно только при $r \geqslant m$, величина U может быть представлена в виде про-

изведения m независимых случайных величин $X_1, \ldots, X_m$:

$$U = \prod_{l=1}^{m} X_l,$$

где X_l имеет β-распределение с параметрами $(n-r-l+1)/2$, $(r-p)/2$ [1].

В соответствии с этими распределениями величины U и решается задача проверки гипотезы $g''=0$. При этом точное определение соответствующих уровней значимости в общем случае невозможно. Методы точного решения этой задачи в частных случаях, а также приближенные методы читатель может найти в [1].

Практически задача проверки гипотезы $g''=0$ в этом случае может быть решена путем оценки вероятности значения U, меньшего, чем полученное экспериментально значение u, методом статистического моделирования. Для этого необходимо моделировать независимые случайные величины с соответствующими β-распределениями и найти частоту события $U < u$ *). Если эта частота достаточно велика, то гипотеза $g''=0$ принимается. В противном случае она отвергается.

Совершенно так же проверяется гипотеза $g''=g''_0$ при любой данной матрице g''_0. Для этого достаточно заменить Y величиной $Y - g''_0 x''$.

Положив $p=0$, можно определить доверительную область для матрицы g, соответствующую коэффициенту доверия α, неравенством $U > u_\alpha$, где u_α определяется из уравнения $P(U > u_\alpha) = \alpha$. Это неравенство определяет случайное множество значений g, которое с вероятностью α накрывает неизвестное значение матрицы g.

Если x — значение некоторой случайной величины X, то в соответствии со сказанным в п. 3.10 легко проверяются только гипотезы о каждом элементе матрицы g по отдельности. В частности, если значение g_{10} коэффициента регрессии g_1 попадает в доверительный интервал

$$\hat{G}_1 - \frac{t_\alpha}{S_{11}} \sqrt{\frac{S_{11}S_{22} - S_{12}^2}{n-r-1}} < g_1 < \hat{G}_1 + \frac{t_\alpha}{S_{11}} \sqrt{\frac{S_{11}S_{22} - S_{12}^2}{n-r-1}},$$

соответствующий коэффициенту доверия $\alpha = P(|T| < t_\alpha)$, то гипотезу $g_1 = g_{10}$ можно принять. В противном случае она отвергается.

---

*) Для вычисления функции распределения в случае β-распределения в [94] служит программа BDTR.

Пример 9. В условиях примера 6 значение $g=0$ не принадлежит найденному доверительному интервалу (0,27, 0,87). Поэтому гипотезу $g=0$ следует отвергнуть.

4.5. Проверка гипотез о линейности регрессии. Вторая гипотеза, которую приходится проверять при построении регрессионных моделей, — это гипотеза о линейности регрессии. Для проверки этой гипотезы необходимо при каждом значении x наблюдать случайную величину Y достаточно большое число раз. Предположим, что величина Y наблюдается n_1 раз при $x=x^{(1)}$, n_2 раз при $x=x^{(2)}$, и т. д., n_N раз при $x=x^{(N)}$, $n_1+\ldots+n_N=N$. Значения величины Y при $x=x^{(p)}$ обозначим $Y_k^{(p)}$ ($k=1, \ldots, n_p$; $p=1, \ldots, n$).

Рассмотрим сначала случай скалярной величины Y. Совершенно так же, как при проверке гипотезы о независимости Y от части координат вектора x, образуем две суммы квадратов отклонений экспериментальных точек от регрессии, одну, S, при гипотезе о линейности регрессии:

$$S=\sum_{p=1}^{N}\sum_{k=1}^{n_p}(Y_k^{(p)}-\hat{G}x^{(p)})^2,$$

и другую, S_0, без этого допущения:

$$S_0=\sum_{p=1}^{N}\sum_{k=1}^{n_p}(Y_k^{(p)}-\bar{Y}^{(p)})^2, \qquad \bar{Y}^{(p)}=\frac{1}{n_p}\sum_{k=1}^{n_p}Y_k^{(p)}.$$

Если отношение $U=S_0/S$ (которое, очевидно, не может быть больше 1) близко к 1, то гипотезу о линейности регрессии можно считать не противоречащей экспериментальным данным. При малом значении U эту гипотезу следует отвергнуть.

В случае векторной величины Y за критерий проверки гипотезы о линейности регрессии можно принять отношение $U=|S_0|/|S|$ определителя матрицы

$$S_0=\sum_{p=1}^{N}\sum_{k=1}^{n_p}(Y_k^{(p)}-\bar{Y}^{(p)})(Y_k^{(p)}-\bar{Y}^{(p)})^T$$

к определителю матрицы

$$S=\sum_{p=1}^{N}\sum_{k=1}^{n_p}(Y_k^{(p)}-\hat{G}x^{(p)})(Y_k^{(p)}-\hat{G}x^{(p)})^T.$$

Совершенно так же, как в предыдущем случае, доказывается, что при нормальном распределении величины Y разность $S-S_0$ независима от S_0 и в случае скалярной Y величины S, S_0 и $S-S_0$ имеют χ^2-распределения соответсвенно с $n-N$, $n-r$ и $N-r$ степенями свободы. Вследствие этого случайная величина U имеет β-распределение с параметрами $(N-r)/2$, $(n-N)/2$ и, соответственно, величина

$$F=\frac{(1-U)/(N-r)}{U/(n-N)}$$

имеет F-распределение $f_{N-r,\,n-N}(f)$ (пример 5.26). С помощью этих распределений гипотеза о линейности регрессии проверяется совершенно так же, как гипотеза $g''=0$ в случае линейной регрессии [24].

Точно так же проверка гипотезы о линейности регрессии в случае векторной величины Y сводится к применению тех же распределений, что и проверка гипотезы $g''=0$ при линейной регрессии.

4.6. Выбор типа регрессионной модели. Рассмотрим теперь вопрос о выборе регрессионной модели. Решение этого вопроса сводится к определению переменных, от которых зависит интересующая нас случайная величина Y, и конечного числа функций этих переменных, которые следует включить в состав вектора x. При решении этой задачи приходится искать разумный компромисс между двумя противоречивыми тенденциями. С одной стороны, естественно стремление учесть возможно больше факторов с целью получить модель, наиболее полно отражающую изучаемое явление. С другой стороны, всегда желательно, чтобы модель была достаточно простой. Практическое применение чрезмерно сложной модели связано с необходимостью получения большого объема информации (приходится измерять и определять каждый раз значения всех переменных, включенных в модель) и выполнения сложных вычислений, требующих больших затрат машинного времени, что в конечном счете приводит к высокой стоимости и трудоемкости применения модели. Кроме того, полнота и точность сложной модели могут оказаться иллюзорными из-за ограниченного объема доступной информации, которой приходится довольствоваться при построении модели. Однако применение чрезмерно простых моделей может приводить к большим ошибкам. Поэтому необходимы модели разумной степени

сложности. Как всегда в подобных случаях, компромиссное решение не может быть однозначным. Более того, две различные модели, полученные в результате компромиссного решения задачи разными методами, трудно сравнить между собой, и невозможно отдать предпочтение какой-нибудь одной из них. Поэтому мы не будем говорить о наилучшей модели и будем называть модель, полученную в результате компромиссного решения, подходящей моделью, подразумевая, что она достаточно хорошо удовлетворяет как требованию полноты описания явления, так и требованию простоты.

Экспериментальным исследованиям, проводимым с целью построения модели, должно предшествовать определение полного набора переменных, которые могут быть включены в модель, и разделение их на входные переменные (которые войдут в состав вектора x и от которых будут зависеть координаты вектора x) и выходные переменные (координаты случайного вектора Y). Лишь после этого могут быть проведены опыты, в результате которых будут получены значения входных и выходных переменных.

Обычно модель строится для каждой из выходных переменных (координат вектора Y) по отдельности, а затем полученные скалярные модели объединяются в одну векторную.

Для построения подходящей регрессионной модели по результатам опытов можно применить следующий прием.

Сначала выявляется характер зависимости выходной переменной от каждой из входных переменных путем грубого глазомерного построения соответствующих графиков. В результате определяются функции входных переменных, которые целесообразно включить в состав вектора x. При этом будут приближенно учтены основные нелинейности модели относительно входных переменных (см. замечание в конце п. 2.7).

После вычисления полученных в результате опыта значений координат вектора x производится вычисление выборочных коэффициентов корреляции всех координат вектора x с выходной переменной Y. Из координат вектора x выбирается одна, для которой коэффициент корреляции с Y имеет наибольшее по модулю значение, и строится модель, включающая только одну эту координату. На этом заканчивается первый шаг. Последующие шаги производятся по такой же схеме.

На каждом шаге вычисляются выборочные коэффициенты корреляции всех еще не включенных в модель координат вектора x с остатком, т. е. с разностью между величиной Y и ее оценкой $\hat{Y}$ с помощью построенной на предыдущем шаге модели. Из этих координат вектора x выбирается та, которой соответствует наибольший по модулю коэффициент корреляции. После этого строится новая регрессионная модель, включающая выбранную таким путем координату, и проверяются гипотезы о возможности пренебречь зависимостью от каждой из ранее включенных в модель координат. В результате принимается окончательная регрессионная модель на этом шаге.

Процесс заканчивается и принимается окончательная регрессионная модель, когда будут перебраны все координаты вектора x или же когда включение новых координат будет приводить к увеличению оценки дисперсии остатка, т. е. к увеличению среднего квадрата ошибки оценивания величины Y с помощью построенной модели.

Изложенный метод построения подходящей регрессионной модели представляет собой один из вариантов шагового регрессионного анализа. Обычно он дает хорошие результаты [24]*).

Пример 10. В условиях примера 4 изложенный метод последовательно включает вторую, пятую и третью координаты вектора x и дает в результате модель примера 8. Попытки включить первую и четвертую координаты приводят к увеличению оценки дисперсии величины Y, т. е. к ухудшению точности модели.

§ 5. Дисперсионный анализ

5.1. Качественные переменные — факторы. В теории, изложенной в §§ 1—4, координаты вектора x могут быть как непрерывно изменяющимися переменными, так и переменными с конечным множеством возможных значений. Однако в некоторых задачах практики наблюдаемая случайная величина Y зависит от факторов, которые невозможно охарактеризовать набором измеряемых скалярных величин, а можно лишь описать словесно. Классическим примером такого рода является исследование зависимости урожая от характера почвы и способа ее обработки (например, вида применяемого удобрения). В данном случае наблюдаемой случайной вели-

*) Другой вариант шагового регрессионного анализа реализуется в стандартной программе STPRG [94].

чиной является вес урожая, который зависит от двух факторов — характера почвы и способа ее обработки. Задача исследования состоит в построении статистической модели этой зависимости. Методы решения подобных задач составляют содержание *дисперсионного анализа* и *теории планирования экспериментов*, основы которых заложены Фишером [112, 116].

Основная задача дисперсионного анализа состоит в том, чтобы по результатам наблюдений некоторой случайной величины Y оценить зависимость ее математического ожидания от факторов, не поддающихся количественному описанию, и ее дисперсию (ковариационную матрицу в случае векторной величины Y). Конкретные значения каждого фактора обычно называют *уровнями* этого фактора. Заметим, что в задачах дисперсионного анализа некоторые факторы могут допускать и количественную характеристику, но при этом опыты производятся только при нескольких фиксированных значениях этих факторов.

5.2. Полный двухфакторный план экспериментов. Рассмотрим сначала случай двух факторов F_1 и F_2. Предположим, что наблюдение величины Y производится при μ уровнях фактора F_1 и ν уровнях фактора F_2, причем для каждой из $\mu\nu$ комбинаций уровней этих факторов Y наблюдается один раз. Такая схема наблюдений называется *полным двухфакторным планом эксперимента*.

Пусть Y_{ij} — значение величины Y, получаемое при i-м уровне F_1 и j-м уровне F_2 ($i=1, \ldots, \mu$; $j=1, \ldots, \nu$). Предположим, что математическое ожидание случайной величины Y_{ij} представимо в виде

$$MY_{ij}=b_0+b_{1i}+b_{2j} \qquad (i=1, \ldots, \mu; \; j=1, \ldots, \nu), \quad (56)$$

где b_0 — общая для всех уровней обоих факторов часть математического ожидания величины Y, b_{1i} и b_{2j} — части, обусловленные влиянием i-го уровня фактора F_1 и j-го уровня фактора F_2 соответственно. Кроме того, предположим, что ковариационная матрица K_y величины Y не зависит от уровней факторов F_1 и F_2. Соотношение (56) выражает обычное для дисперсионного анализа предположение об аддитивности влияния двух факторов.

Легко сообразить, что параметр b_0 можно выбрать произвольно, в частности, принять равным 0. Однако практически удобно принять b_0 равным среднему значению MY для всех возможных комбинаций уровней F_1

и F_2. При этом b_{1i} и b_{2j} будут связаны соотношениями

$$\sum_{i=1}^{\mu} b_{1i} = 0, \qquad \sum_{j=1}^{\nu} b_{2j} = 0. \tag{57}$$

Задача построения модели в данном случае сводится к оцениванию величин b_0, b_{1i} $(i = 1, \ldots, \mu)$ и b_{2j} $(j = 1, \ldots, \nu)$ и матрицы K_y по результатам наблюдений.

5.3. Приведение к задаче оценивания линейной регрессии. Заметим, что величины MY_{ij} $(i = 1, \ldots, \mu;$ $j = 1, \ldots, \nu)$ представляют собой значения функции

$$f(x) = b_0 x_1 + b_{11} x_2 + \ldots + b_{1,\,\mu-1} x_\mu + \\ + b_{21} x_{\mu+1} + \ldots + b_{2,\,\nu-1} x_{\mu+\nu-1}$$

$\mu + \nu - 1$ переменных, из которых $x_1 = 1$ всегда, а остальные принимают одно из значений 0, 1, —1.

▶ Действительно, в силу (57) при $x_{i+1} = x_{\mu+j} = 1$, $x_h = 0$ $(h = 2, \ldots, \mu + \nu - 1;\ h \neq i + 1,\ \mu + j)$

$$f(x) = b_0 + b_{1i} + b_{2j} = MY_{ij}$$
$$(i = 1, \ldots, \mu - 1;\ j = 1, \ldots, \nu - 1),$$

при $x_{i+1} = 1$, $x_h = 0$ $(h = 2, \ldots, \mu - 1;\ h \neq i + 1)$, $x_h = -1$ $(h = \mu + 1, \ldots, \mu + \nu - 1)$

$$f(x) = b_0 + b_{1i} - b_{21} - \ldots - b_{2,\,\nu-1} = b_0 + b_i + b_{2\nu} = MY_{i\nu}$$
$$(i = 1, \ldots, \mu - 1),$$

при $x_{\mu+j} = 1$, $x_h = 0$ $(h = \mu + 1, \ldots, \mu + \nu - 1;\ h \neq \mu + j)$, $x_{i+1} = -1$ $(i = 1, \ldots, \mu - 1)$

$$f(x) = b_0 - b_{11} - \ldots - b_{1,\,\mu-1} + b_{2j} = b_0 + b_{1\mu} + b_{2j} = MY_{\mu j}$$
$$(j = 1, \ldots, \nu - 1),$$

при $x_2 = \ldots = x_{\mu+\nu-1} = -1$

$$f(x) = b_0 - b_{11} - \ldots - b_{1,\,\mu-1} - b_{21} - \ldots - b_{2,\,\nu-1} = \\ = b_0 + b_{1\mu} + b_{2\nu} = MY_{\mu\nu}.$$ ◀

Таким образом, задача дисперсионного анализа сводится к задаче линейного регрессионного анализа § 3 при

$$gx = b_0 x_1 + b_{11} x_2 + \ldots + b_{1,\,\mu-1} x_\mu + b_{21} x_{\mu+1} + \ldots \\ \ldots + b_{2,\,\nu-1} x_{\mu+\nu-1}.$$

При этом координаты вектора x (за исключением $x_1 = 1$) могут иметь только три возможных значения 0, 1 и —1. Следовательно, формулы для оценок параметров b_0, b_{1i},

b_{2j} и K_y и доверительных областей для них можно вывести из общих формул § 3. В результате получим оптимальные несмещенные оценки $\hat{B}_0$, $\hat{B}_{1i}$ и $\hat{B}_{2j}$ параметров b_0, b_{1i} и b_{2j}:

$$\hat{B}_0 = \frac{1}{\mu\nu} \sum_{i=1}^{\mu} \sum_{j=1}^{\nu} Y_{ij},$$

$$\hat{B}_{1i} = \frac{1}{\nu} \sum_{j=1}^{\nu} Y_{ij} - \hat{B}_0 \quad (i = 1, \ldots, \mu), \tag{58}$$

$$\hat{B}_{2j} = \frac{1}{\mu} \sum_{i=1}^{\mu} Y_{ij} - \hat{B}_0 \quad (j = 1, \ldots, \nu).$$

Несмещенная оценка ковариационной матрицы K_y получается на основании (37) в результате деления статистики

$$S = \sum_{i=1}^{\mu} \sum_{j=1}^{\nu} (Y_{ij} - \hat{B}_0 - \hat{B}_{1i} - \hat{B}_{2j})(Y_{ij} - \hat{B}_0 - \hat{B}_{1i} - \hat{B}_{2j})^T$$

на число опытов $\mu\nu$ за вычетом размерности $\mu + \nu - 1$ вектора x, т. е. на $(\mu - 1)(\nu - 1)$:

$$\hat{K}_y = \frac{1}{(\mu-1)(\nu-1)} \sum_{i=1}^{\mu} \sum_{j=1}^{\nu} (Y_{ij} - \hat{B}_0 - \hat{B}_{1i} - \hat{B}_{2j}) \times$$
$$\times (Y_{ij} - \hat{B}_0 - \hat{B}_{1i} - \hat{B}_{2j})^T. \tag{59}$$

Матрица A в данном случае имеет блочно-диагональную структуру:

$$A = \begin{bmatrix} \mu\nu & 0 & 0 \\ 0 & A_1 & 0 \\ 0 & 0 & A_2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 2\nu & \nu & \ldots & \nu \\ \nu & 2\nu & \ldots & \nu \\ \cdot & \cdot & \cdot\cdot\cdot & \cdot \\ \nu & \nu & \ldots & 2\nu \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 2\mu & \mu & \ldots & \mu \\ \mu & 2\mu & \ldots & \mu \\ \cdot & \cdot & \cdot\cdot\cdot & \cdot \\ \mu & \mu & \ldots & 2\mu \end{bmatrix},$$

где блоки с нулями представляют собой матрицы, все элементы которых равны 0, матрица A_1 имеет размер $(\mu - 1) \times (\mu - 1)$, а $A_2 - (\nu - 1) \times (\nu - 1)$. Поэтому на основании (32) случайный вектор $\hat{B}_0$ и случайные матрицы $\hat{B}_1 = [\hat{B}_{11} \ldots \hat{B}_{1\mu}]$ и $\hat{B}_2 = [\hat{B}_{21} \ldots \hat{B}_{2\nu}]$ не коррелированы.

Из (32), принимая во внимание, что

$$A_1^{-1}=\frac{1}{\mu\nu}\begin{bmatrix}\mu-1 & -1 & \dots & -1\\ -1 & \mu-1 & \dots & -1\\ \dots & \dots & \dots & \dots\\ -1 & -1 & \dots & \mu-1\end{bmatrix},$$

$$A_2^{-1}=\frac{1}{\mu\nu}\begin{bmatrix}\nu-1 & -1 & \dots & -1\\ -1 & \nu-1 & \dots & -1\\ \dots & \dots & \dots & \dots\\ -1 & -1 & \dots & \nu-1\end{bmatrix},$$

получаем формулы для ковариационной матрицы K_0 вектора $\hat{B}_0$:

$$K_0=K_y/\mu\nu, \tag{60}$$

для ковариационных и взаимных ковариационных матриц $K_{ih}^{(1)}$ векторов $\hat{B}_{11}, \dots, \hat{B}_{1,\ \mu-1}$:

$$K_{ii}^{(1)}=\frac{\mu-1}{\mu\nu}K_y,\quad K_{ih}^{(1)}=-\frac{K_y}{\mu\nu}\qquad (i,\ h=1,\ \dots,\ \mu-1;\ h\neq i) \tag{61}$$

и для ковариационных и взаимных ковариационных матриц $K_{lj}^{(2)}$ векторов $\hat{B}_{21}, \dots, \hat{B}_{2,\ \nu-1}$:

$$K_{jj}^{(2)}=\frac{\nu-1}{\mu\nu}K_y,\quad K_{lj}^{(2)}=-\frac{K_y}{\mu\nu}\qquad (j,\ l=1,\ \dots,\ \nu-1;\ l\neq j). \tag{62}$$

Эти формулы распространяются по симметрии и на векторы $\hat{B}_{1\mu}=-\hat{B}_{11}-\dots-\hat{B}_{1,\ \mu-1}$ и $\hat{B}_{2\nu}=-B_{21}-\dots-\hat{B}_{2,\ \nu-1}$ — несмещенные оценки параметров $b_{1\mu}$ и $b_{2\nu}$. Несмещенные оценки матриц K_0, $K_{ih}^{(1)}$ и $K_{lj}^{(2)}$ получаются заменой в (60) — (62) матрицы K_y ее несмещенной оценкой $\hat{K}_y$ из (59).

Найдем еще оценки параметров

$$K_1=\frac{1}{\mu-1}\sum_{i=1}^{\mu}b_{1i}b_{1i}^T,\qquad K_2=\frac{1}{\nu-1}\sum_{j=1}^{\nu}b_{2j}b_{2j}^T, \tag{63}$$

обычно называемых *ковариациями средних за счет факторов* F_1 *и* F_2 соответственно. Естественно принять за оценки K_1 и K_2 соответствующие статистические средние

$$K_1^*=\frac{1}{\mu-1}\sum_{i=1}^{\mu}\hat{B}_{1i}\hat{B}_{1i}^T,\quad K_2^*=\frac{1}{\nu-1}\sum_{j=1}^{\nu}\hat{B}_{2j}\hat{B}_{2j}^T.$$

Однако эти оценки смещенные. Чтобы получить несмещенные оценки K_1 и K_2, найдем математическое ожидание K_1^*. Имея в виду, что $\hat{B}_{1i}$ — несмещенные оценки параметров b_{1i} и поэтому связаны теми же соотношениями (57), что и b_{1i}, в силу (61) имеем

$$M\sum_{i=1}^{\mu}\hat{B}_{1i}\hat{B}_{1i}^T = M\sum_{i=1}^{\mu}(\hat{B}_{1i}-b_{1i})(\hat{B}_{1i}-b_{1i})^T + \sum_{i=1}^{\mu}b_{1i}b_{1i}^T =$$

$$= \mu\frac{\mu-1}{\mu\nu}K_y + \sum_{i=1}^{\mu}b_{1i}b_{1i}^T = \frac{\mu-1}{\nu}K_y + \sum_{i=1}^{\mu}b_{1i}b_{1i}^T.$$

Поэтому

$$MK_1^* = K_1 + \frac{1}{\nu}K_y.$$

Отсюда видно, что для устранения смещения оценки K_1^* достаточно вычесть из нее величину $\hat{K}_y/\nu$. В результате, учитывая (59), получим несмещенную оценку K_1:

$$\hat{K}_1 = \frac{1}{\mu-1}\left[\sum_{i=1}^{\mu}\hat{B}_{1i}\hat{B}_{1i}^T - \frac{S}{\nu(\nu-1)}\right]. \qquad (64)$$

По симметрии получаем несмещенную оценку параметра K_2:

$$\hat{K}_2 = \frac{1}{\nu-1}\left[\sum_{j=1}^{\nu}\hat{B}_{2j}\hat{B}_{2j}^T - \frac{S}{\mu(\mu-1)}\right]. \qquad (65)$$

В случае нормального распределения величины Y из некоррелированности оценок $\hat{B}_0$, $\hat{B}_1$ и $\hat{B}_2$ вытекает их независимость (§ 3.5). Отсюда и из доказанного в § 3 следует, что статистики

$$S_0 = \mu\nu(\hat{B}_0-b_0)(\hat{B}_0-b_0)^T,$$

$$S_1 = (\hat{B}_1-b_1)A_1(\hat{B}_1-b_1)^T = \nu\sum_{i=1}^{\mu}(\hat{B}_{1i}-b_{1i})(\hat{B}_{1i}-b_{1i})^T,$$

$$S_2 = (\hat{B}_2-b_2)A_2(\hat{B}_2-b_2)^T = \mu\sum_{j=1}^{\nu}(\hat{B}_{2j}-b_{2j})(\hat{B}_{2j}-b_{2j})^T,$$

$$S = \sum_{i=1}^{\mu}\sum_{j=1}^{\nu}(Y_{ij}-\hat{B}_0-\hat{B}_{1i}-\hat{B}_{2j})(Y_{ij}-\hat{B}_0-\hat{B}_{1i}-\hat{B}_{2j})^T$$

независимы и имеют распределения Уишарта $w_{2m}(s)$, $w_{\mu m}(s)$, $w_{\nu m}(s)$ и $w_{(\mu-1)(\nu-1)+1,\, m}(s)$ соответственно. В част-

ном случае скалярной величины Y величины S_0/D_y, S_1/D_y, S_2/D_y и S/D_y имеют χ^2-распределения с 1, $\mu-1$, $\nu-1$ и $(\mu-1)(\nu-1)$ степенями свободы соответственно.

В дополнение к общим способам построения доверительных областей для b_0, b_1, b_2 и K_y и проверки гипотез о b_0, b_1, b_2 §§ 3 и 4 в данном случае возникают дополнительные возможности вследствие независимости S_0, S_1 и S_2. В частности, в случае скалярной величины Y для нахождения доверительных областей для b_0, b_1 и b_2 по отдельности и для проверки гипотез о них можно воспользоваться тем, что отношения

$$(\mu-1)(\nu-1)\,S_0/S, \quad (\nu-1)\,S_1/S, \quad (\mu-1)\,S_2/S$$

имеют F-распределения $f_{1,\,(\mu-1)(\nu-1)}(f)$, $f_{\mu-1,\,(\mu-1)(\nu-1)}(f)$ и $f_{\nu-1,\,(\mu-1)(\nu-1)}(f)$ соответственно. Это дает для матриц b_1 и b_2 доверительные области

$$\sum_{i=1}^{\mu} (\hat{B}_{1i}-b_{1i})^2 < f_\alpha \frac{S}{\nu(\nu-1)},$$

$$\sum_{j=1}^{\nu} (\hat{B}_{2j}-b_{2j})^2 < f_\alpha \frac{S}{\mu(\mu-1)}, \tag{66}$$

где f_α — верхняя $100(1-\alpha)$-процентная точка F-распределения $f_{\mu-1,\,(\mu-1)(\nu-1)}$ или $f_{\nu-1,\,(\mu-1)(\nu-1)}$ соответственно. Для проверки гипотезы о равенстве b_{1i} и b_{2j} данным значениям (обычно 0) достаточно установить, удовлетворяют эти значения соответствующему неравенству (66) или нет. В первом случае гипотеза принимается, а во втором отвергается.

Совершенно так же дисперсионный анализ производится при любом числе факторов. При этом в случае *полного N-факторного плана*, когда величина Y наблюдается при всех комбинациях уровней факторов, кроме влияния факторов по отдельности, можно учесть еще взаимодействие их по 2, по 3, ..., по $N-1$. Так, например, в случае полного трехфакторного плана можно предположить, что

$$MY_{ijk} = b_0 + b_{1i} + b_{2j} + b_{3k} + b_{12,\,ij} + b_{13,\,ik} + b_{23,\,jk}.$$

Пример 11 *). Даны результаты полного четырехфакторного эксперимента при четырех уровнях первого фактора, трех — второго и третьего и двух — четвертого:

j		1				2				3			
l	k \ i	1	2	3	4	1	2	3	4	1	2	3	4
1	1	3	10	9	8	24	8	9	3	2	8	9	8
	2	4	12	3	9	22	7	16	2	2	2	7	2
	3	5	10	5	8	23	9	17	3	2	8	6	3
2	1	2	14	9	13	29	16	11	3	2	7	5	3
	2	7	11	5	8	28	18	10	6	6	6	5	9
	3	9	10	27	8	28	16	11	7	8	9	8	15

Учитывая взаимодействия всех факторов по два и по три, строим модель вида

$$MY_{ijkl} = b_0 + b_{1i} + b_{2j} + b_{3k} + b_{4l} + b_{12ij} + b_{13ik} + b_{14il} + \\ + b_{23jk} + b_{24jl} + b_{34kl} + b_{123ijk} + b_{124ijl} + b_{134ikl} + b_{234jkl}, \qquad (I)$$

где $b_{1i}, \ldots, b_{12ij}, \ldots, b_{123ijk}, \ldots$ удовлетворяют условиям типа (57):

$$\sum_i b_{1i} = \sum_j b_{2j} = \sum_k b_{3k} = \sum_l b_{4l} = 0,$$

$$\sum_i b_{12ij} = \sum_j b_{12ij} = \ldots = \sum_k b_{34kl} = \sum_l b_{34kl} = 0,$$

$$\sum_i b_{123ijk} = \sum_j b_{123ijk} = \sum_k b_{123ijk} = \ldots = \sum_l b_{234jkl} = 0.$$

Число независимых параметров $b_0, b_{1i}, \ldots, b_{234jkl}$ $(i = 1, 2, 3, 4;$ $j, k = 1, 2, 3;$ $l = 1, 2)$ $r = 60$. Оценки параметров $b_0, b_{1i}, \ldots, b_{12ij}, \ldots$

*) Исходные данные этого примера и результаты вычисления на ЭВМ сумм квадратов по программе ANOVA заимствованы из [94].

Программа дисперсионного анализа ANOVA включает три подпрограммы: AVDAT, AVCAL, MEANQ. Эта программа реализует вычисления для дисперсионного анализа факторных планов с числом факторов до 6, а при соответствующей модификации и с большим числом факторов [94].

$\ldots, b_{123ijk}, \ldots$ определяются формулами

$$\hat{B}_0 = \overline{Y} = \frac{1}{72} \sum_{i,\, j,\, k,\, l} Y_{ijkl}, \qquad \hat{B}_{1i} = \frac{1}{18} \sum_{j,\, k,\, l} Y_{ijkl} - \hat{B}_0, \ \ldots,$$

$$\hat{B}_{12ij} = \frac{1}{6} \sum_{k,\, l} Y_{ijkl} - \hat{B}_0 - \hat{B}_{1i} - \hat{B}_{2j}, \ \ldots,$$

$$\hat{B}_{123ijk} = \frac{1}{2} \sum_{l} Y_{ijkl} - \hat{B}_0 - \hat{B}_{1i} - \hat{B}_{2j} - \hat{B}_{3k} - \hat{B}_{12ij} - \hat{B}_{13ik} - \hat{B}_{23jk}, \ldots$$

Суммы

$$S = \sum_{i,\, j,\, k,\, l} (Y_{ijkl} - \hat{B}_0 - \hat{B}_{1i} - \hat{B}_{2j} - \hat{B}_{3k} - \hat{B}_{4l} - \hat{B}_{12ij} - \hat{B}_{13ik} -$$
$$- \hat{B}_{14il} - \hat{B}_{23jk} - \hat{B}_{24jl} - \hat{B}_{34kl} - \hat{B}_{123ijk} - \hat{B}_{124ijl} - \hat{B}_{134ikl} - \hat{B}_{234jkl})^2,$$

$$S_1 = 18 \sum_{i} (\hat{B}_{1i} - b_{1i})^2, \qquad S_2 = 24 \sum_{j} (\hat{B}_{2j} - b_{2j})^2,$$

$$S_3 = 24 \sum_{k} (\hat{B}_{3k} - b_{3k})^2, \qquad S_4 = 36 \sum_{l} (\hat{B}_{4l} - b_{4l})^2,$$

$$S_{12} = 6 \sum_{i,\, j} (\hat{B}_{12ij} - b_{12ij})^2, \qquad S_{13} = 6 \sum_{i,\, k} (\hat{B}_{13ik} - b_{13ik})^2,$$

$$S_{14} = 9 \sum_{i,\, l} (\hat{B}_{14il} - b_{14il})^2, \qquad S_{23} = 8 \sum_{j,\, k} (\hat{B}_{23jk} - b_{23jk})^2,$$

$$S_{24} = 12 \sum_{j,\, l} (\hat{B}_{24jl} - b_{24jl})^2, \qquad S_{34} = 12 \sum_{k,\, l} (\hat{B}_{34kl} - b_{34kl})^2,$$

$$S_{123} = 2 \sum_{i,\, j,\, k} (\hat{B}_{123ijk} - b_{123ijk})^2, \qquad S_{124} = 3 \sum_{i,\, j,\, l} (\hat{B}_{124ijl} - b_{124ijl})^2,$$

$$S_{134} = 3 \sum_{i,\, k,\, l} (\hat{B}_{134ikl} - b_{134ikl})^2, \qquad S_{234} = 4 \sum_{j,\, k,\, l} (\hat{B}_{234jkl} - b_{234jkl})^2$$

в случае нормального распределения величины Y с независимой от факторов (постоянной) дисперсией D_y независимы и имеют χ^2-распределения соответственно с $k = 12$, $k_1 = 3$, $k_2 = k_3 = 2$, $k_4 = 1$, $k_{12} = k_{13} = 6$, $k_{14} = 3$, $k_{23} = 4$, $k_{24} = k_{34} = 2$, $k_{123} = 12$, $k_{124} = k_{134} = 6$ и $k_{234} = 4$ степенями свободы.

Результаты вычислений.

1) Оценки параметров $b_0, b_{1i}, \ldots, b_{234jkl}$: $\hat{b}_0 = 9{,}40278$, а остальные оценки даны на стр. 389 и 390.

2) Значения сумм $S, S_1, \ldots, S_{234}$ при $b_{1i} = \ldots = b_{234jkl} = 0$ и оценка дисперсии величины Y: $s = 151{,}02777$, $s_1 = 229{,}04166$, $s_2 = 722{,}69434$, $s_3 = 55{,}11110$, $s_4 = 141{,}68054$, $s_{12} = 1382{,}08325$, $s_{13} = 42{,}00000$, $s_{14} = 18{,}81944$, $s_{23} = 13{,}13889$, $s_{24} = 6{,}02778$, $s_{34} = 40{,}77777$, $s_{123} = 140{,}75000$, $s_{124} = 176{,}97221$, $s_{134} = 50{,}55554$, $s_{234} = 62{,}63889$, $\hat{D}_y = 151{,}02777/12 = 12{,}58566$.

3) Доверительные области для параметров $b_{1i}, \ldots, b_{234jkl}$, соответствующие коэффициенту доверия $\alpha = 0{,}95$, при нормальном распределении величины Y (с учетом результата примера 5.24 относительно

i, j, k, l	$\hat{b}_{1i}$	$\hat{b}_{2j}$	$\hat{b}_{3k}$	$\hat{b}_{4l}$
1	2,041	−0,695	−0,444	−1,403
2	0,653	4,180	−0,778	1,402
3	0,153	−3,486	1,222	
4	−2,847			

	$\hat{b}_{12ij}$			$\hat{b}_{13ik}$			$\hat{b}_{14il}$	
j, k, l	1	2	3	1	2	3	1	2
i								
1	−5,749	10,043	−4,291	−0,667	0,834	−0,166	−0,374	0,376
2	1,806	−1,903	0,096	0,888	0,055	−0,946	−0,431	0,431
3	0,806	−1,403	0,597	−0,445	−1,111	1,555	0,847	−0,847
4	3,139	−6,736	3,597	0,221	0,222	−0,445	−0,042	0,042

	$\hat{b}_{23jk}$				$\hat{b}_{24jl}$			$\hat{b}_{34kl}$	
k	1	2	3	l	1	2	l	1	2
j				j			k		
1	0,236	−0,555	0,320	1	−0,138	0,140	1	0,861	−0,861
2	−0,264	0,820	−0,555	2	−0,263	0,265	2	0,111	−0,110
3	0,027	−0,261	0,236	3	0,403	−0,402	3	−0,972	0,972

$\hat{b}_{123ij}$

k	1			2			3		
j	1	2	3	1	2	3	1	2	3
i									
1	−1,625	2,208	−0,583	0,999	−1,543	0,538	0,623	−0,668	0,041
2	0,153	−0,513	0,363	1,611	0,070	−1,682	−1,763	0,446	1,312
3	−0,014	−1,180	1,195	−3,223	1,735	1,488	3,236	−0,555	−2,680
4	1,487	−0,513	−0,972	0,611	−0,264	−0,347	−2,097	0,778	1,320

$\hat{b}_{124ijl}$						
l	1			2		
j	1	2	3	1	2	3
i						
1	0,915	—0,627	—0,293	—0,918	0,625	0,290
2	1,472	—2,236	0,765	—1,473	2,236	—0,764
3	—3,306	2,486	0,819	3,305	—2,486	—0,820
4	0,916	0,375	—1,292	—0,917	—0,375	—1,291

$\hat{b}_{134ikl}$						
l	1			2		
k	1	2	3	1	2	3
i						
1	0,250	—0,501	0,249	—0,250	0,499	—0,250
2	—0,860	—0,610	1,474	0,861	0,611	—1,470
3	0,028	1,445	—1,472	—0,028	—1,445	1,473
4	0,584	—0,333	—0,249	—0,583	0,333	0,251

$\hat{b}_{234jkl}$						
l	1			2		
k	1	2	3	1	2	3
j						
1	—0,320	1,055	—0,737	0,319	—1,057	0,736
2	—1,070	—0,320	1,388	1,068	0,318	—1,389
3	1,389	—0,729	—0,653	—1,389	0,732	0,653

распределения отношений $(S_1/3)/(S/12)$, ...):

$$18\,[(2{,}041-b_{11})^2+(0{,}653-b_{12})^2+(0{,}153-b_{23})^2+(2{,}847+b_{14})^2] <$$

$$< \frac{3}{12} f_\alpha \cdot 151{,}02777 = \frac{3}{12} \cdot 3{,}49 \cdot 151{,}02777 = 130{,}963,$$

. . . .

$$12\left[(0{,}861-b_{34,\,11})^2+(0{,}111-b_{34,\,21})^2+(0{,}972+b_{34,\,31})^2+\right.$$
$$\left.+(0{,}861+b_{3412})^2+(0{,}110+b_{3422})^2+(0{,}972-b_{3432})^2\right]<$$
$$<\frac{2}{12}f_\alpha\cdot 151{,}02777=\frac{2}{12}\cdot 3{,}88\cdot 151{,}02777=97{,}670,$$

. . . .

4) Проверка гипотез о независимости величины Y от факторов.

Нулевые значения параметров b_{3k}, b_{13ik}, b_{14il}, b_{23jk}, b_{24jl}, b_{34kl}, b_{123ijk}, b_{124ijl}, b_{134ikl}, b_{234jkl} попадают в соответствующие доверительные области (в этом легко убедиться, подставив полученные значения сумм s_3, s_{13}, s_{14}, s_{23}, s_{24}, s_{34}, s_{123}, s_{124}, s_{134}, s_{234} при нулевых значениях параметров b_{3k}, b_{13ik}, b_{14il}, b_{23jk}, b_{24jl}, b_{34kl}, b_{123ijk}, b_{124ijl}, b_{134ikl}, b_{234jkl} в левые части неравенств, определяющих доверительные области). Поэтому можно принять гипотезу о независимости Y от третьего фактора, взятого отдельно, и в комбинациях с любыми другими факторами и об аддитивной зависимости Y от четвертого фактора (т. е. об отсутствии взаимодействия четвертого фактора с первым и вторым). В результате принимаем упрощенную модель

$$MY_{ijkl}=b_0+b_{1i}+b_{2j}+b_{4l}+b_{12ij}. \qquad \text{(II)}$$

При этом новая сумма S будет представлять собой сумму $S+S_3+$ $+S_{13}+S_{14}+S_{23}+S_{24}+S_{34}+S_{123}+S_{124}+S_{134}$ и ее χ^2-распределение будет иметь 59 степеней свободы (число независимых параметров b_0, b_{1i}, b_{2j}, b_{4l}, b_{12ij} $(i=1, 2, 3, 4;\ \ j=1, 2, 3;\ \ l=1, 2)$ $r=13$). Оценка дисперсии величины Y будет равна

$$\hat{D}_y=\frac{757{,}8239}{59}=12{,}84447,$$

а доверительные области для параметров b_{1i}, b_{2j}, b_{4l} и b_{12ij} определятся неравенствами

$$18\left[(2{,}041-b_{11})^2+(0{,}653-b_{12})^2+(0{,}153-b_{13})^2+(2{,}847+b_{14})^2\right]<$$
$$<\frac{3}{59}f_\alpha\cdot 757{,}8239=\frac{3}{59}\cdot 3{,}15\cdot 757{,}8239=121{,}380,$$

. . . .

Полученные результаты иллюстрируют сказанное в п. 4.6 относительно того, что более сложная модель может оказаться менее точной, чем простая, при данном объеме информации. В данном случае простая модель (II) практически не хуже сложной модели (I) по точности, так как оценки дисперсии величины Y (остаточной дисперсии в случае детерминированной модели) у них практически одинаковы, а доверительные области для параметров у модели (II) даже несколько меньше, чем у модели (I).

5.4. Неполные планы экспериментов. Полные многофакторные планы редко удается осуществить из-за практической невозможности провести наблюдения при всех возможных комбинациях уровней факторов. Поэтому для практики особенно важное значение имеют неполные планы экспериментов. При этом возникают задачи опти-

мального выбора комбинаций уровней факторов при данном числе экспериментов с целью получения минимальных дисперсий оценок параметров. Методы решения таких задач дает теория планирования экспериментов [7—9, 38, 63, 108, 116].

Простейшим неполным трехфакторным планом эксперимента является так называемый *план латинского квадрата*. В этом плане все три фактора имеют одно и то же число уровней μ и при каждой комбинации уровней любых двух факторов наблюдение производится только для одного уровня третьего фактора. При этом число наблюдений равно μ^2 вместо μ^3 в случае полного трехфакторного плана.

Ясно, что для неполных планов не все взаимодействия факторов могут быть оценены из-за недостаточности числа наблюдений. Так, например, в плане латинского квадрата взаимодействия факторов невозможно оценить, так как для каждой комбинации уровней двух факторов имеется лишь одно наблюдение.

Дисперсионный анализ для неполных планов производится совершенно так же, как для полных. Только суммирование при вычислении средних производится не по всем возможным комбинациям уровней факторов, а только по тем, для которых производятся наблюдения. Соответственно изменяются знаменатели и числа слагаемых в выражениях средних.

Принимая для плана латинского квадрата

$$MY_{ijk}=b_0+b_{1i}+b_{2j}+b_{3k} \qquad (i,\ j,\ k=1,\ \ldots,\ \mu),$$

$$\sum_{i=1}^{\mu} b_{1i}=\sum_{j=1}^{\mu} b_{2j}=\sum_{k=1}^{\mu} b_{3k}=0,$$

получим несмещенные оценки параметров b_0, b_{1i}, b_{2j}, b_{3k} и K_y:

$$\hat{B}_0=\frac{1}{\mu^2}\sum_{I} Y_{ijk}, \quad \hat{B}_{1i}=\frac{1}{\mu}\sum_{I_1(i)} Y_{ijk}-\hat{B}_0,$$

$$\hat{B}_{2j}=\frac{1}{\mu}\sum_{I_2(j)} Y_{ijk}-\hat{B}_0, \quad \hat{B}_{3k}=\frac{1}{\mu}\sum_{I_3(k)} Y_{ijk}-\hat{B}_0, \tag{67}$$

$$\hat{K}_y=\frac{S}{(\mu-1)(\mu-2)},$$

$$S=\sum_{I}(Y_{ijk}-\hat{B}_0-\hat{B}_{1i}-\hat{B}_{2j}-\hat{B}_{3k}) \times \tag{68}$$

$$\times (Y_{ijk}-\hat{B}_0-\hat{B}_{1i}-\hat{B}_{2j}-\hat{B}_{3k})^T,$$

где I — множество комбинаций индексов i, j, k, для которых производятся наблюдения, $I_1(i)$ — множество комбинаций индексов j, k, для которых при данном i производятся наблюдения, $I_1(i)=\{(j, k): (i, j, k)\in I\}$, и точно так же определяются $I_2(j)=\{(i, k): (i, j, k)\in I\}$ и $I_3(k)=\{(i, j):(i, j, k)\in I\}$. Вектор $\hat{B}_0$ и матрицы $\hat{B}_p = [\hat{B}_{p1}\ldots\hat{B}_{p\mu}]$ $(p=1, 2, 3)$ не коррелированы, и ковариационные матрицы векторов $\hat{B}_0$, $\hat{B}_{1i}$, $\hat{B}_{2j}$ и $\hat{B}_{3k}$ определяются формулами (60)—(62) при $\nu=\mu$. Несмещенные оценки этих матриц получаются заменой в (60)—(62) матрицы K_y ее несмещенной оценкой из (68). Несмещенные оценки ковариаций средних за счет каждого из трех факторов

$$K_p=\frac{1}{\mu-1}\sum_{i=1}^{\mu} b_{pi}b_{pi}^T \qquad (p=1, 2, 3)$$

определяются формулой

$$\hat{K}_p=\frac{1}{\mu-1}\left[\sum_{i=1}^{\mu}\hat{B}_{pi}\hat{B}_{pi}^T-\frac{S}{\mu(\mu-2)}\right] \qquad (p=1, 2, 3). \quad (69)$$

В соответствии с блочно-диагональной структурой матрицы A § 3 в данном случае величина $(\hat{G}-g)\,A\,(\hat{G}-g)^T$ представляет собой сумму четырех слагаемых:

$$S_0=\mu^2(\hat{B}_0-b_0)(\hat{B}_0-b_0)^T,$$
$$S_p=\mu\sum_{i=1}^{\mu}(\hat{B}_{pi}-b_{pi})(\hat{B}_{pi}-b_{pi})^T \qquad (p=1, 2, 3). \quad (70)$$

В случае нормального распределения Y величины $\hat{B}_0$, $\hat{B}_1$, $\hat{B}_2$, $\hat{B}_3$ и S независимы. Вследствие этого статистики S_0, S_1, S_2, S_3 и S независимы и имеют распределения Уишарта $w_{2m}(s)$, $w_{\mu m}(s)$ (общее распределение S_1, S_2 и S_3) и $w_{(\mu-1)(\mu-2)+1,\,m}(s)$. Если Y — скалярная величина, то S_0/D_y, S_1/D_y, S_2/D_y, S_3/D_y и S/D_y имеют χ^2-распределения с 1, $\mu-1$ и $(\mu-1)(\mu-2)$ степенями свободы. В соответствии с этим определяются доверительные области для $\hat{B}_0$, $\hat{B}_1$, $\hat{B}_2$ и $\hat{B}_3$ и проверяются гипотезы о них по отдельности и в любых комбинациях.

Более подробное изложение методов дисперсионного анализа читатель найдет в [123].

Глава 10

СТАТИСТИЧЕСКИЕ МОДЕЛИ, II

§ 1. Модели, описываемые разностными уравнениями

1.1. Модели авторегрессии. В приложениях часто встречаются последовательности случайных величин, определяемые рекуррентным соотношением вида

$$Y_{n+p}=\varphi_n(Y_n, \ldots, Y_{n+p-1})+\psi_n(Y_n, \ldots, Y_{n+p-1})V_n, \quad (1)$$

где $\{\varphi_n\}$ и $\{\psi_n\}$ — некоторые последовательности функций, а $\{V_n\}$ — последовательность независимых случайных величин с нулевыми математическими ожиданиями *). Соотношение (1) определяет модель формирования последовательности $\{Y_n\}$ из последовательности независимых случайных величин $\{V_n\}$. Такие модели называются *моделями авторегрессии* вследствие того, что первое слагаемое в правой части (1) представляет собой условное математическое ожидание величины Y_{n+p} относительно $Y_n, \ldots, Y_{n+p-1}$, т. е. регрессию каждого члена Y_{n+p} последовательности $\{Y_n\}$ на непосредственно предшествующие члены той же последовательности $Y_n, \ldots, Y_{n+p-1}$. Число p называется *порядком* модели авторегрессии.

Последовательность $\{Y_n\}$ может быть конечной ($n=1, \ldots, N$), или односторонне бесконечной ($n=1, 2, \ldots$), или двусторонне бесконечной ($n=0, \pm 1, \pm 2, \ldots$). В первых двух случаях уравнение (1) определяет члены последовательности $\{Y_n\}$, начиная лишь с Y_{p+1}, а $Y_1, \ldots, Y_p$ могут быть произвольными случайными величинами (конечно, независимыми от величин V_n).

Уравнение (1) представляет собой разностное уравнение p-го порядка, определяющее последовательность

*) Предположение, что $MV_n=0$, очевидно, не ограничивает общности, так как если $MV_n\neq 0$, то величину $\psi(Y_n, \ldots, Y_{n+p-1})MV_n$ всегда можно включить в функцию $\varphi(Y_n, \ldots, Y_{n+p-1})$. Тогда во втором слагаемом останется центрированная случайная величина $V_n^0=V_n-MV_n$.

$\{Y_n\}$, так как $Y_{n+1}, \ldots, Y_{n+p}$ можно выразить через Y_n и разности $\Delta Y_n = Y_{n+1} - Y_n, \ldots, \Delta^p Y_n = \Delta^{p-1} Y_{n+1} - \Delta^{p-1} Y_n$.

1.2. Линейные модели. В случае, когда функции φ_n линейны, а ψ_n не зависят от $Y_n, \ldots, Y_{n+p-1}$, уравнение (1) определяет *линейную модель авторегрессии* *):

$$Y_{n+p} = \sum_{r=0}^{p-1} a_{rn} Y_{n+r} + V_n,$$

где $a_{0n}, a_{1n}, \ldots, a_{p-1,n}$ — некоторые коэффициенты (матрицы в случае векторных величин Y_n).

Кроме моделей авторегрессии, часто встречаются модели вида

$$Y_n = \sum_{s=0}^{p} a_{sn} V_{n-s}.$$

Такие модели называются *моделями скользящего среднего*.

Ясно, что линейные модели авторегрессии и модели скользящего среднего представляют собой частные случаи *комбинированной модели*, определяемой уравнением

$$Y_{n+p} = \sum_{r=0}^{p-1} a_{rn} Y_{n+r} + \sum_{s=0}^{p} b_{sn} V_{n+s}. \tag{2}$$

Обратим внимание на то, что во всех рассмотренных моделях величина Y_{n+p} не зависит от последующих величин Y_k и V_l. Это соответствует реальным ситуациям, когда любая система может использовать лишь сигналы, поступившие на нее в прошлом, но не может реагировать на сигналы, которые подействуют на нее в будущем (условие *физической возможности* системы).

1.3. Приведение линейной модели к модели авторегрессии первого порядка. Докажем, что комбинированную модель (в частности, линейную модель авторегрессии) любого порядка можно привести к линейной модели авторегрессии первого порядка.

▶ Введем новые случайные величины $Z_{1n}, \ldots, Z_{pn}$, определив их соотношениями

$$\begin{aligned} &Y_n = Z_{1n} + q_{1n} V_n, \\ &Z_{k,n+1} = Z_{k+1,n} + q_{k+1,n} V_n \quad (k = 1, \ldots, p-1), \end{aligned} \tag{3}$$

где $q_{1n}, \ldots, q_{pn}$ — пока неизвестные коэффициенты (в общем случае матрицы). Тогда будем иметь

$$Y_{n+1} = Z_{1,n+1} + q_{1,n+1} V_{n+1} = Z_{2n} + q_{2n} V_n + q_{1,n+1} V_{n+1}$$

*) В этом случае постоянные множители ψ_n можно включить в V_n.

и вообще

$$Y_{n+r}=Z_{r+1,\,n}+q_{r+1,\,n}V_n+q_{r,\,n+1}V_{n+1}+\dots+q_{1,\,n+r}V_{n+r}$$

$$(r=1,\ \dots,\ p-1). \tag{4}$$

Заменив в этой формуле при $r=p-1$ n на $n+1$, получим

$$Y_{n+p}=Z_{p,\,n+1}+q_{p,\,n+1}V_{n+1}+\dots+q_{1,\,n+p}V_{n+p}=$$

$$=Z_{p,\,n+1}+\sum_{s=1}^{p}q_{p-s+1,\,n+s}V_{n+s}. \tag{5}$$

Подставив выражения (4) и (5) в (2), будем иметь

$$Z_{p,\,n+1}+\sum_{s=1}^{p}q_{p-s+1,\,n+s}V_{n+s}=$$

$$=\sum_{r=0}^{p-1}a_{rn}\left(Z_{r+1,\,n}+q_{r+1,\,n}V_n+\dots+q_{1,\,n+r}V_{n+r}\right)+$$

$$+\sum_{s=0}^{p}b_{sn}V_{n+s},$$

или

$$Z_{p,\,n+1}+\sum_{s=1}^{p}q_{p-s+1,\,n+s}V_{n+s}=$$

$$=\sum_{r=0}^{p-1}a_{rn}Z_{r+1,\,n}+\sum_{s=0}^{p-1}V_{n+s}\sum_{r=s}^{p-1}a_{rn}q_{r-s+1,\,n+s}+\sum_{s=0}^{p}b_{sn}V_{n+s}. \tag{6}$$

Определим теперь $q_{1n},\ \dots,\ q_{pn}$ так, чтобы уравнение (6) не содержало $V_{n+1},\ \dots,\ V_{n+p}$. Сравнивая коэффициенты при V_{n+p} в левой и правой частях уравнения (6), получаем

$$q_{1,\,n+p}=b_{pn}.$$

Отсюда, заменив n на $n-p$, находим

$$q_{1n}=b_{p,\,n-p}.$$

Сравнивая коэффициенты при V_{n+s} $(s=1,\ \dots,\ p-1)$ в левой и правой частях уравнения (6), получим

$$q_{p-s+1,\,n+s}=\sum_{r=s}^{p-1}a_{rn}q_{r-s+1,\,n+s}+b_{sn}.$$

Отсюда, заменив n на $n-s$, находим

$$q_{p-s+1,\,n}=b_{s,\,n-s}+\sum_{r=s}^{p-1}a_{r,\,n-s}\,q_{r-s+1,\,n}.$$

Введем теперь вместо индексов s и r новые индексы k и l, положив $p-s+1=k$, $r-s+1=l$. Тогда будем иметь $s=p-k+1$, $r=p-k+l$, и полученная формула примет вид

$$q_{kn}=b_{p-k+1,\, n-p+k-1}+\sum_{l=1}^{k-1} a_{p-k+l,\, n-p+k-1}\, q_{ln}$$
$$(k=2, \ldots, p).$$

Определив таким образом $q_{1n}, \ldots, q_{pn}$, найдем коэффициент $q_{p+1,\, n}$ при V_n в (6):

$$q_{p+1,\, n}=b_{0n}+\sum_{r=0}^{p-1} a_{rn} q_{r+1,\, n}.$$

В результате уравнение (6) примет вид

$$Z_{p,\, n+1}=\sum_{r=0}^{p-1} a_{rn} Z_{r+1,\, n}+q_{p+1,\, n} V_n. \tag{7}$$

Равенства (3) и (7) представляют собой систему разностных уравнений первого порядка, определяющую величины $Z_{1n}, \ldots, Z_{pn}$. Положим теперь

$$Z_n=\begin{bmatrix} Z_{1n} \\ Z_{2n} \\ \ldots \\ Z_{pn} \end{bmatrix}, \quad c_n=\begin{bmatrix} 0 & I & 0 & \ldots & 0 \\ 0 & 0 & I & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & I \\ a_{0n} & a_{1n} & a_{2n} & \ldots & a_{p-1,\, n} \end{bmatrix},$$

$$W_n=\begin{bmatrix} q_{2n} \\ q_{3n} \\ \ldots \\ q_{p+1,\, n} \end{bmatrix} V_n,$$

где 0 — квадратная матрица той же размерности, что и $a_{0n}, a_{1n}, \ldots, a_{n-1,\, n}$, все элементы которой равны нулю, а I — единичная матрица той же размерности. Тогда уравнения (3) и (7) запишутся в виде одного уравнения

$$Z_{n+1}=c_n Z_n+W_n.$$

Таким образом, мы получили линейную регрессионную модель первого порядка. ◀

1.4. Нелинейные модели. Кроме линейных и нелинейных моделей авторегрессии, на практике встречаются и более общие модели, описываемые нелинейными разностными уравнениями вида

$$Y_{n+1}=\omega_n(Y_n,\, V_n), \tag{8}$$

где $\{\omega_n\}$ — последовательность известных функций, а $\{V_n\}$ — последовательность независимых случайных величин.

К модели (8) приводится и более общая модель, описываемая уравнением

$$Y_{n+p} = \omega_n(Y_n, \ldots, Y_{n+p-1}, V_n).$$

Для этого достаточно ввести последовательность векторов $\{Z_n\}$, $Z_n = [Y_n^T \ldots Y_{n+p-1}^T]^T$, и добавить тождественные уравнения $Y_{n+1} = Y_{n+1}, \ldots, Y_{n+p-1} = Y_{n+p-1}$, которые можно записать в виде $Z_{n+1,1} = Z_{n2}, \ldots, Z_{n+1, p-1} = Z_{np}$. Эти уравнения вместе с данным уравнением можно записать в виде уравнения (8) для последовательности составных векторов $\{Z_n\}$.

В частности, уравнение авторегрессии (1) приводится таким способом к уравнению (8) при $\omega_n(y, v) = \varphi_n(y) + \psi_n(y)v$.

Построение моделей авторегрессии, а также некоторых моделей типа (8) по экспериментальным данным производится методами, изложенными в §§ 9.3 и 9.4 (см. также [34]).

§ 2. Оценивание величин, определяемых разностным уравнением

2.1. Нелинейные модели в общем случае. Большое значение для практики имеет задача оценивания последовательности случайных векторов $\{Y_n\}$, определяемых разностным уравнением вида (8), по результатам измерения на каждом шаге некоторых функций вектора Y_n.

Так как измерения всегда сопровождаются ошибками, то полученная в результате измерений последовательность случайных векторов $\{X_n\}$ в общем случае определяется формулой

$$X_n = \omega_n'(Y_n, V_n), \tag{9}$$

где $\{\omega_n'\}$ — последовательность известных функций *).

---

*) В задачах практики ошибки измерений обычно не зависят от случайных величин V_n в уравнении (8) формирования последовательности векторов $\{Y_n\}$. Вследствие этого правая часть уравнения (9) представляет собой функцию $\omega_n'(Y_n, U_n)$, где $\{U_n\}$ — последовательность независимых случайных величин, независимая от $\{V_n\}$. Однако $\omega_n(Y_n, V_n)$ и $\omega_n'(Y_n, U_n)$ можно рассматривать как функции Y_n и составного вектора $W_n = [U_n^T \, V_n^T]^T$. Поэтому случай независимых ошибок измерений и случайных величин в уравнении формирования векторов Y_n можно рассматривать как частный слу-

Таким образом, возникает задача оценивания случайных векторов Y_n, определяемых разностным уравнением (8), по результатам наблюдения случайных векторов X_n, определяемых формулой (9).

Будем искать оптимальную оценку $\hat{Y}_{n+1}$ каждой величины Y_{n+1} в классе функций величин $X_1, \ldots, X_n$, определяемых разностными уравнениями

$$\hat{Y}_{n+1} = \delta_n \zeta_n (\hat{Y}_n, X_n) + \gamma_n \tag{10}$$

при заданной последовательности функций $\{\zeta_n\}$ и при всех возможных значениях δ_n и γ_n. При этом будем предполагать, что распределения всех случайных величин V_n известны. В качестве функций ζ_n в (10) можно взять векторные функции любой размерности. Соответственно δ_n будут прямоугольными матрицами, а γ_n — векторами той же размерности, что Y_n и $\hat{Y}_n$.

Поставленная задача сводится к определению оптимальных значений δ_n и γ_n при каждом n, при которых средний квадрат ошибки $M|\hat{Y}_{n+1} - Y_{n+1}|^2$ минимален.

На основании сказанного в § 9.2 задача определения оптимальных значений δ_n и γ_n в (10) равноценна задаче нахождения линейной средней квадратической регрессии случайного вектора Y_{n+1} на случайный вектор $X'_n = \zeta_n (Y_n, X_n)$. Следовательно, оптимальные значения δ_n и γ_n определяются уравнением (9.14) и формулой (9.15), которые в данном случае имеют вид

$$\delta_n K_n = L_n, \qquad \gamma_n = m_n - \delta_n l_n, \tag{11}$$

где K_n — ковариационная матрица случайного вектора $X'_n = \zeta_n (\hat{Y}_n, X_n)$ *),

$$K_n = M[\zeta_n (\hat{Y}_n, X_n) - M\zeta_n (\hat{Y}_n, X_n)]\, \zeta_n (\hat{Y}_n, X_n)^T,$$

L_n — взаимная ковариационная матрица векторов Y_{n+1} и $X'_n = \zeta_n (\hat{Y}_n, X_n)$,

$$L_n = M(Y_{n+1} - MY_{n+1})\, \zeta_n (\hat{Y}_n, X_n)^T,$$

чай, когда координаты каждого случайного вектора V_n распадаются на две независимые группы, от одной из которых зависит функция ω_n в (8), а от другой — функция ω'_n в (9).

*) Для любых случайных векторов Z и U $K_{zu} = MZ^0 U^{0T} = MZ^0 (U^T - m_u^T) = MZ^0 U^T$.

а m_n и l_n — математические ожидания векторов Y_{n+1} и $X'_n = \zeta_n(\hat{Y}_n, X_n)$ соответственно,

$$m_n = MY_{n+1}, \quad l_n = M\zeta_n(\hat{Y}_n, X_n).$$

Соотношение (9.10) и вытекающее из него соотношение $M(\hat{Y} - Y)\hat{Y}^* = 0$ в данном случае дают

$$M(\hat{Y}_{n+1} - Y_{n+1})\zeta_n(\hat{Y}_n, X_n)^T = 0, \quad M(\hat{Y}_{n+1} - Y_{n+1})\hat{Y}^T_{n+1} = 0. \tag{12}$$

Для определения K_n, L_n, m_n и l_n достаточно знать совместное распределение величин Y_n и $\hat{Y}_n$ при каждом n. Действительно, подставив в формулы для K_n, L_n, m_n и l_n выражения величин Y_{n+1} и X_n из уравнений (8) и (9), получим K_n, L_n, m_n и l_n как математические ожидания некоторых функций случайных величин Y_n, $\hat{Y}_n$ и V_n. Но Y_n и $\hat{Y}_n$ являются определенными функциями величин $V_1, \ldots, V_{n-1}$ в силу уравнений (8)—(10) и, следовательно, не зависят от V_n. Поэтому для вычисления K_n, L_n, m_n и l_n достаточно знать совместное распределение величин Y_n и $\hat{Y}_n$ и распределение величины V_n, которое по предположению известно.

Чтобы определить совместное распределение величин Y_n и $\hat{Y}_n$, достаточно найти их характеристическую функцию

$$g_n(\lambda, \mu) = M\exp\{i\lambda^T Y_n + i\mu^T \hat{Y}_n\}.$$

Для этого подставим сюда выражения Y_n и $\hat{Y}_n$ из уравнений (8) и (10) с заменой n на $n-1$ и после этого подставим в полученное выражение величину X_{n-1} из уравнения (9), заменив в нем n на $n-1$. В результате получим

$$g_n(\lambda, \mu) = M\exp\{i\lambda^T\omega_{n-1}(Y_{n-1}, V_{n-1}) + \\ + i\mu^T[\delta_{n-1}\zeta_{n-1}(\hat{Y}_{n-1}, \omega'_{n-1}(Y_{n-1}, V_{n-1})) + \gamma_{n-1}]\}. \tag{13}$$

Правая часть этой формулы представляет собой математическое ожидание известной функции величин Y_{n-1}, $\hat{Y}_{n-1}$ и V_{n-1}, причем Y_{n-1} и $\hat{Y}_{n-1}$ в силу (8)—(10) независимы от V_{n-1}. Поэтому для вычисления характеристической функции $g_n(\lambda, \mu)$ величин Y_n и $\hat{Y}_n$ достаточно знать характеристическую функцию $g_{n-1}(\lambda, \mu)$ величин Y_{n-1} и $\hat{Y}_{n-1}$ и

распределение величины V_{n-1}, которое по предположению известно.

Таким образом, формулы (11) и (13) определяют рекуррентный процесс, позволяющий находить на каждом шаге g_n, δ_n и γ_n по известным g_{n-1}, δ_{n-1} и γ_{n-1}.

Чтобы начать рекуррентный процесс, необходимо задать совместное распределение величин Y_1 и $\hat{Y}_1$ (начальное распределение).

Если распределение величины Y_1 (независимой от V_1) известно и определяется характеристической функцией $k_1(\lambda)$, можно принять за оценку $\hat{Y}_1$ математическое ожидание m_0 величины Y_1, $\hat{Y}_1 = m_0 = MY_1$, и определить совместную характеристическую функцию Y_1 и $\hat{Y}_1$ формулой

$$g_1(\lambda, \mu) = k_1(\lambda) e^{i\mu^T m_0}.$$

Можно также определить оценку $\hat{Y}_1$ как случайную величину с выбранным произвольно распределением и принять

$$g_1(\lambda, \mu) = k_1(\lambda) l_1(\mu),$$

где $l_1(\mu)$ — характеристическая функция величины $\hat{Y}_1$.

Если распределение величины Y_1 не известно, то его приходится задавать произвольно. После этого начальная характеристическая функция $g_1(\lambda, \mu)$ определится так же, как и в предыдущем случае.

Так как величины X_n не входят в уравнения (11) и (13), определяющие δ_n, γ_n и g_n, то все последовательности величин $\{g_n\}$, $\{\delta_n\}$ и $\{\gamma_n\}$ можно вычислить изложенным методом заранее. Тогда процесс оценивания последовательности величин $\{Y_n\}$ сведется к применению рекуррентной формулы (10).

Изложенный метод оценивания случайных величин Y_n, определяемых разностным уравнением (8), в силу свойств линейной оценки регрессии со сдвигом дает несмещенную оценку $\hat{Y}_n$ каждой из величин Y_n при $n > 1$ (оценка $\hat{Y}_1$ выбирается произвольно, вследствие чего она может быть смещенной).

Точность оценки $\hat{Y}_n$ можно характеризовать вторым моментом ошибки

$$R_n = M(\hat{Y}_n - Y_n)(\hat{Y}_n^T - Y_n^T)$$

или средним квадратом ошибки

$$\varepsilon_n = M(\hat{Y}_n^T - Y_n^T)(\hat{Y}_n - Y_n) = \operatorname{tr} R_n,$$

которые определяются характеристической функцией $g_n(\lambda, \mu)$. Можно также находить доверительные области для Y_n, поскольку $g_n(\lambda, \mu)$ полностью определяет совместное распределение величин Y_n и $\hat{Y}_n$.

Изложенный метод дает условно оптимальные (оптимальные в данном классе функций) оценки величин Y_n при любых функциях ζ_n. Однако качество получаемых оценок зависит от того, как выбраны функции ζ_n. Естественно рассчитывать на то, что достаточно хорошие оценки можно получить, если выбрать функции ζ_n так, чтобы уравнение (10), определяющее оценки, было близким по форме к уравнению (8), определяющему оцениваемые случайные величины. Других рекомендаций о выборе функций ζ_n пока еще дать нельзя. Вопрос о рациональном выборе ζ_n нуждается в дальнейших исследованиях [81, 82].

Пример 1. Последовательность скалярных случайных величин $\{Y_n\}$ и последовательность наблюдаемых случайных величин $\{X_n\}$ определяются уравнениями

$$Y_{n+1} = (Y_n + V_n)^2, \qquad X_n = Y_n + U_n,$$

где $\{V_n\}$ и $\{U_n\}$ — независимые последовательности независимых случайных величин с нулевыми моментами первого и третьего порядков. Найти оптимальные рекуррентные оценки величин Y_n, удовлетворяющие разностному уравнению первого порядка

$$\hat{Y}_{n+1} = \delta_n [\hat{Y}_n^2 \;\; \hat{Y}_n(X_n - \hat{Y}_n) \;\; (X_n - \hat{Y}_n)^2]^T + \gamma_n.$$

Выбор в качестве $\delta_n \zeta_n$ квадратичной формы переменных $\hat{Y}_n$ и $X_n - \hat{Y}_n$ в данном случае естествен, так как при этом уравнение для оценок получается похожим на уравнение, определяющее величины Y_n, вследствие чего можно надеяться получить достаточно точные оценки величин Y_n.

Оптимальные коэффициенты δ_n и γ_n определяются в данном случае формулами (11). При этом получаем следующие формулы для элементов матриц $L_n = [l_1^{(n)} \;\; l_2^{(n)} \;\; l_3^{(n)}]$ и $K_n = [k_{ij}^{(n)}]$ и для математических ожиданий m_n и l_n:

$$l_1^{(n)} = MY_n^2\hat{Y}_n^2 - MY_n^2 M\hat{Y}_n^2,$$

$$l_2^{(n)} = MY_n^3\hat{Y}_n - MY_n^2 MY_n\hat{Y}_n - l_1^{(n)},$$

$$l_3^{(n)} = MY_n^4 - (MY_n^2)^2 - l_1^{(n)} - 2l_2^{(n)},$$

$$k_{11}^{(n)} = M\hat{Y}_n^4 - (M\hat{Y}^2)^2,$$

$$k_{12}^{(n)} = MY_n\hat{Y}_n^3 - MY_n\hat{Y}_n M\hat{Y}_n^2 - k_{11}^{(n)},$$

$$k_{13}^{(n)} = l_1^{(n)} - k_{11}^{(n)} - 2k_{12}^{(n)},$$

$$k_{22}^{(n)} = MY_n^2\hat{Y}_n^2 - (MY_n\hat{Y}_n)^2 + H_n M\hat{Y}_n^2 - k_{11}^{(n)} - 2k_{12}^{(n)},$$

$$k_{23}^{(n)}=l_1^{(n)}+l_2^{(n)}-k_{11}^{(n)}-3k_{12}^{(n)}-k_{13}^{(n)}-2k_{22}^{(n)}+2H_nMY_n\hat{Y}_n,$$
$$k_{33}^{(n)}=l_1^{(n)}+2l_2^{(n)}+l_3^{(n)}-k_{11}^{(n)}-4k_{12}^{(n)}-2k_{13}^{(n)}-4k_{22}^{(n)}-4k_{23}^{(n)}+H_n'-H_n^2+$$
$$+4H_nMY_n^2,$$
$$m_n=MY_n^2+G_n,$$
$$l_n=[M\hat{Y}_n^2\ MY_n\hat{Y}_n-M\hat{Y}_n^2\ MY_n^2+M\hat{Y}_n^2-2MY_n\hat{Y}_n+H_n],$$

где H_n и G_n — дисперсии величин U_n и V_n соответственно, а H_n' — четвертый момент величины U_n. В случае нормально распределенных величин U_n $H_n'=3H_n^2$. Для вычисления вторых и четвертых моментов величин Y_n и $\hat{Y}_n$ в полученных формулах достаточно знать совместное распределение величин Y_n и $\hat{Y}_n$. При этом моменты можно вычислять или дифференцированием характеристической функции $g_n(\lambda, \mu)$ величин Y_n и $\hat{Y}_n$, или интегрированием, считая известной плотность

$$f_n(y, \eta)=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i\lambda y-i\mu\eta}g_n(\lambda, \mu)\,d\lambda\,d\mu$$

величин Y_n и $\hat{Y}_n$:

$$MY_n^2=-\left[\frac{\partial^2 g_n}{\partial\lambda^2}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f_n(y, \eta)\,dy\,d\eta,$$

$$MY_n\hat{Y}_n=-\left[\frac{\partial^2 g_n}{\partial\lambda\,\partial\mu}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\eta f_n(y, \eta)\,dy\,d\eta,$$

$$M\hat{Y}_n^2=-\left[\frac{\partial^2 g_n}{\partial\mu^2}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \eta^2 f_n(y, \eta)\,dy\,d\eta,$$

$$MY_n^4=\left[\frac{\partial^4 g_n}{\partial\lambda^4}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^4 f_n(y, \eta)\,dy\,d\eta,$$

$$MY_n^3\hat{Y}_n=\left[\frac{\partial^4 g_n}{\partial\lambda^3\,\partial\mu}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^3\eta f_n(y, \eta)\,dy\,d\eta,$$

$$MY_n^2\hat{Y}_n^2=\left[\frac{\partial^4 g_n}{\partial\lambda^2\,\partial\mu^2}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2\eta^2 f_n(y, \eta)\,dy\,d\eta,$$

$$MY_n\hat{Y}_n^3=\left[\frac{\partial^4 g_n}{\partial\lambda\,\partial\mu^3}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\eta^3 f_n(y, \eta)\,dy\,d\eta,$$

$$M\hat{Y}_n^4=\left[\frac{\partial^4 g_n}{\partial\mu^4}\right]_0=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \eta^4 f_n(y, \eta)\,dy\,d\eta,$$

где индексом 0 внизу отмечены значения величин в квадратных скобках при $\lambda = \mu = 0$.

Уравнение (13) для характеристической функции $g_n(\lambda, \mu)$ имеет в данном случае вид

$$g_n(\lambda, \mu) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{i\lambda (y + v)^2 + i\mu (\delta_{n-1} [\eta^2 \ \eta (y - \eta + u) \ (y - \eta + u)^2]^T + \gamma_{n-1})\} \times f_{n-1}(y, \eta)\, p_{n-1}(v)\, q_{n-1}(u)\, du\, dv\, dy\, d\eta,$$

где $p_{n-1}(v)$ и $q_{n-1}(u)$ — плотности величин V_{n-1} и U_{n-1} соответственно.

Полученные формулы определяют рекуррентный процесс для последовательного вычисления g_n, f_n, δ_n и γ_n.

Аналогично решается задача экстраполяции (прогноза) последовательности $\{Y_n\}$, т. е. оценивания вектора Y_{n+p} после получения наблюдения величины X_n ($p > 1$). Для экстраполяции на $p - 1$ шагов следует определить оценки $\hat{Y}_n$ векторов Y_n уравнением

$$\hat{Y}_{n+p} = \delta_n \zeta_n (\hat{Y}_n, \ldots, \hat{Y}_{n+p-1}, X_n) + \gamma_n.$$

При этом все предыдущие выкладки останутся практически неизменными. Только для нахождения K_n, L_n, m_n и l_n в (11) необходимо будет определить совместную характеристическую функцию $g_n(\lambda, \mu_1, \ldots, \mu_p)$ случайных векторов Y_n, $\hat{Y}_n$, $\ldots$, $\hat{Y}_{n+p-1}$. Эта характеристическая функция определяется совершенно так же, как $g_n(\lambda, \mu)$ в рассмотренном случае [81, 82].

Заметим, что изложенный метод дает также возможность оценивать не все координаты векторов Y_n, а только часть их. Для этого достаточно взять функции $\zeta_n(y, x)$ (или $\zeta_n(y_1, \ldots, y_p, x)$ в случае экстраполяции), зависящие только от соответствующей части координат вектора y ($y_1, \ldots, y_p$ в случае экстраполяции).

2.2. Нелинейные модели авторегрессии. Рассмотрим частный случай нелинейных уравнений авторегрессии, когда уравнения (8) и (9) линейны относительно величины V_n. В этом случае

$$\omega_n(y, v) = \varphi_n(y) + \psi_n(y) v, \quad \omega'_n(y, v) = \varphi'_n(y) + \psi'_n(y) v, \quad (14)$$

где φ_n, ψ_n, φ'_n и ψ'_n — известные функции. В этом случае естественно выбрать функции ζ_n в (10) так, чтобы урав-

нение (10) было линейным относительно X_n, т. е. принять

$$\zeta_n(y, x) = [\xi_n(y)^T \; x^T \eta_n(y)]^T \tag{15}$$

и соответственно положить $\delta_n = [\alpha_n \; \beta_n]$.

Уравнения (8) и (9) имеют в данном случае вид

$$Y_{n+1} = \varphi_n(Y_n) + \psi_n(Y_n) V_n, \tag{16}$$

$$X_n = \varphi_n'(Y_n) + \psi_n'(Y_n) V_n, \tag{17}$$

а уравнение (10) можно записать в виде

$$\hat{Y}_{n+1} = \alpha_n \xi_n(\hat{Y}_n) + \beta_n \eta_n(\hat{Y}_n) X_n + \gamma_n. \tag{18}$$

Матрицы K_n, L_n и векторы m_n и l_n определяются при этом формулами

$$K_n = \begin{bmatrix} \varkappa_{11}^{(n)} & \varkappa_{12}^{(n)} \\ \varkappa_{21}^{(n)} & \varkappa_{22}^{(n)} \end{bmatrix}, \quad L_n = [\varkappa_{01}^{(n)} \; \varkappa_{02}^{(n)}], \quad l_n = \begin{bmatrix} l_1^{(n)} \\ l_2^{(n)} \end{bmatrix},$$

где

$$\varkappa_{01}^{(n)} = M[\varphi_n(Y_n) - M\varphi_n(Y_n)]\,\xi_n(\hat{Y}_n)^T,$$

$$\varkappa_{02}^{(n)} = M[\varphi_n(Y_n) - M\varphi_n(Y_n)]\,\varphi_n'(Y_n)^T \eta_n(\hat{Y}_n)^T + \\ + M\psi_n(Y_n) G_n \psi_n'(Y_n)^T \eta_n(\hat{Y}_n)^T,$$

$$\varkappa_{11}^{(n)} = M[\xi_n(\hat{Y}_n) - M\xi_n(\hat{Y}_n)]\,\xi_n(\hat{Y}_n)^T,$$

$$\varkappa_{12}^{(n)} = \varkappa_{21}^{(n)T} = M[\xi_n(\hat{Y}_n) - M\xi_n(\hat{Y}_n)]\,\varphi_n'(Y_n)^T \eta_n(\hat{Y}_n)^T,$$

$$\varkappa_{22}^{(n)} = M[\eta_n(\hat{Y}_n)\varphi_n'(Y_n) - M\eta_n(\hat{Y}_n)\varphi_n'(Y_n)]\,\varphi_n'(Y_n)^T \eta_n(Y_n)^T + \\ + M\eta_n(\hat{Y}_n)\psi_n'(Y_n) G_n \psi_n'(Y_n)^T \eta_n(\hat{Y}_n)^T,$$

$$m_n = M\varphi_n(Y_n),$$

$$l_1^{(n)} = M\xi_n(\hat{Y}_n), \qquad l_2^{(n)} = M\eta_n(\hat{Y}_n)\varphi_n'(Y_n).$$

Подставив выражения матриц K_n и L_n и $\delta_n = [\alpha_n \; \beta_n]$ в первое уравнение (11), получаем уравнения для оптимальных значений α_n и β_n в (18):

$$\begin{aligned} \alpha_n \varkappa_{11}^{(n)} + \beta_n \varkappa_{21}^{(n)} &= \varkappa_{01}^{(n)}, \\ \alpha_n \varkappa_{12}^{(n)} + \beta_n \varkappa_{22}^{(n)} &= \varkappa_{02}^{(n)}. \end{aligned} \tag{19}$$

Подставив выражение вектора l_n и $\delta_n = [\alpha_n \; \beta_n]$ во вторую формулу (11), получаем формулу для оптимального значения γ_n:

$$\gamma_n = m_n - \alpha_n l_1^{(n)} - \beta_n l_2^{(n)}. \tag{20}$$

Положив в (12) $\zeta_n(\hat{Y}_n, \ X_n)^T = [\xi_n(\hat{Y}_n)^T \ X_n^T \eta_n(\hat{Y}_n)^T]$, получим соотношения

$$M(\hat{Y}_{n+1} - Y_{n+1})\,\xi_n(\hat{Y}_n)^T = 0,$$
$$M(\hat{Y}_{n+1} - Y_{n+1})\,X_n^T \eta_n(\hat{Y}_n)^T = 0. \quad (21)$$

Второе соотношение (12) при этом не изменится.

Подставив выражения (14) и (15) функций ω_n, ω'_n и ζ_n и $\delta_{n-1} = [\alpha_{n-1} \ \beta_{n-1}]$ в уравнение (13), получим

$$g_n(\lambda, \mu) = M \exp\{i\lambda^T[\varphi_{n-1}(Y_{n-1}) + \psi_{n-1}(Y_{n-1})V_{n-1}] + i\mu^T[\alpha_{n-1}\xi_{n-1}(\hat{Y}_{n-1}) + \beta_{n-1}\eta_{n-1}(\hat{Y}_{n-1})\varphi'_{n-1}(Y_{n-1}) + \beta_{n-1}\eta_{n-1}(\hat{Y}_{n-1})\psi'_{n-1}(Y_{n-1})V_{n-1} + \gamma_{n-1}]\}.$$

Но, как уже отмечалось, величины Y_{n-1} и $\hat{Y}_{n-1}$ независимы от V_{n-1}. Следовательно, условное математическое ожидание

$$M[\exp\{i[\lambda^T\psi_{n-1}(Y_{n-1}) + \mu^T\eta_{n-1}(\hat{Y}_{n-1})\psi'_{n-1}(Y_{n-1})]V_{n-1} \mid Y_{n-1}, \ \hat{Y}_{n-1}\}$$

совпадает с безусловным математическим ожиданием, представляющим собой значение характеристической функции $h_{n-1}(\lambda')$ случайной величины V_{n-1} при $\lambda' = \psi_{n-1}(Y_{n-1})^T\lambda + \psi'_{n-1}(Y_{n-1})^T\eta_{n-1}(\hat{Y}_{n-1})^T\mu$. Учитывая это, можем написать уравнение для характеристической функции $g_n(\lambda, \mu)$ в виде

$$g_n(\lambda, \mu) = Mh_{n-1}(\psi_{n-1}(Y_{n-1})^T\lambda + \psi'_{n-1}(Y_{n-1})^T\eta_{n-1} \times (\hat{Y}_{n-1})^T\beta_{n-1}^T\mu)\exp\{i\lambda^T\varphi_{n-1}(Y_{n-1}) + i\mu^T[\alpha_{n-1}\xi_{n-1}(\hat{Y}_{n-1}) + \beta_{n-1}\eta_{n-1}(\hat{Y}_{n-1})\varphi'_{n-1}(Y_{n-1}) + \gamma_{n-1}]\}. \quad (22)$$

Уравнения (19), (20) и (22) определяют рекуррентный процесс для последовательного вычисления g_n, α_n, β_n, γ_n по известным g_{n-1}, α_{n-1}, β_{n-1}, γ_{n-1}. Для начала процесса необходимо задать характеристическую функцию $g_1(\lambda, \mu)$ величин Y_1 и $\hat{Y}_1$. Это делается совершенно так же, как в общем случае п. 2.1.

В случае экстраполяции последовательности $\{Y_n\}$ на $p-1$ шагов уравнение (18) заменяется уравнением

$$\hat{Y}_{n+p} = \alpha_n\xi_n(\hat{Y}_n, \ldots, \hat{Y}_{n+p-1}) + \beta_n\eta_n(\hat{Y}_n, \ldots, \hat{Y}_{n+p-1})X_n + \gamma_n.$$

Для определения коэффициентов $\varkappa_{ij}^{(n)}$ в уравнениях (19) и величин m_n, $l_1^{(n)}$ и $l_2^{(n)}$ в формуле (20) необходимо в

общем случае знать совместную характеристическую функцию $g_n(\lambda, \mu_1, \ldots, \mu_p)$ величин Y_n, $\hat{Y}_n, \ldots, \hat{Y}_{n+p-1}$. Формула для этой характеристической функции выводится совершенно так же, как (22) [81, 82].

Относительно выбора функций ξ_n и η_n можно дословно повторить все сказанное в п. 2.1 о выборе функции ζ_n. Однако в некоторых случаях целесообразно потребовать, чтобы уравнение (18), определяющее оценки, было линейным (линейное оценивание). В таких случаях следует принять $\xi_n(y) = y$ $(\xi_n(y_1, \ldots, y_p) = [y_1^T \ldots y_p^T]^T$ в случае экстраполяции), $\eta_n(y) = I$.

2.3. Линейные модели авторегрессии. Рассмотрим как частный случай задачу линейного оценивания случайных величин, определяемых линейной моделью авторегрессии

$$Y_{n+1} = a_n Y_n + a_{n0} + \psi_n V_n \tag{23}$$

по результатам наблюдения последовательности случайных величин

$$X_n = b_n Y_n + b_{n0} + \psi_n' V_n. \tag{24}$$

В этом случае оценки $\hat{Y}_n$ векторов Y_n определяются уравнением

$$\hat{Y}_{n+1} = \alpha_n \hat{Y}_n + \beta_n X_n + \gamma_n.$$

Для нахождения оптимальных значений α_n, β_n и γ_n надо положить во всех формулах п. 2.2 $\varphi_n(y) = a_n y + a_0$, $\varphi_n'(y) = b_n y + b_0$, $\xi_n(y) = y$, $\eta_n(y) = I$, а ψ_n и ψ_n' считать постоянными. Тогда получим

$$\varkappa_{01}^{(n)} = a_n M(Y_n - MY_n)\hat{Y}_n^T = a_n K_{y_n\hat{y}_n},$$

$$\varkappa_{02}^{(n)} = a_n M(Y_n - MY_n)(Y_n^T b_n^T + b_{n0}^T) + \psi_n G_n \psi_n' = a_n K_{y_n} b_n^T + \psi_n G_n \psi_n',$$

$$\varkappa_{11}^{(n)} = M(\hat{Y}_n - M\hat{Y}_n)\hat{Y}_n^T = K_{\hat{y}_n},$$

$$\varkappa_{12}^{(n)} = \varkappa_{21}^{(n)\,T} = M(\hat{Y}_n - M\hat{Y}_n)(Y_n^T b_n^T + b_{n0}^T) = K_{\hat{y}_n y_n} b_n^T,$$

$$\varkappa_{22}^{(n)} = b_n M(Y_n - MY_n)(Y_n^T b_n^T + b_{n0}^T) + \psi_n' G_n \psi_n'^T = b_n K_{y_n} b_n^T + \psi_n' G_n \psi_n'^T,$$

$$m_n = a_n MY_n + a_{n0}, \quad l_1^{(n)} = M\hat{Y}_n, \quad l_2^{(n)} = b_n MY_n + b_{n0},$$

где K_{y_n}, $K_{\hat{y}_n}$, $K_{y_n\hat{y}_n}$ и $K_{\hat{y}_n y_n}$ — ковариационные и взаимные ковариационные матрицы векторов Y_n и $\hat{Y}_n$. Соотношения (21) в данном случае имеют вид

$$M(\hat{Y}_{n+1} - Y_{n+1})\hat{Y}_n^T = 0, \qquad M(\hat{Y}_{n+1} - Y_{n+1})X_n^T = 0. \tag{25}$$

Второе равенство (12) после замены $n+1$ на n дает

$$M(\hat{Y}_n - Y_n)\hat{Y}_n^T = M(\hat{Y}_n^0 - Y_n^0)\hat{Y}_n^T = K_{\hat{y}_n} - K_{y_n\hat{y}_n} = 0.$$

Следовательно, $K_{y_n\hat{y}_n} = K_{\hat{y}_n}$ и в силу симметрии матрицы $K_{\hat{y}_n}$ $K_{\hat{y}_n y_n} = K_{y_n\hat{y}_n}^T = K_{\hat{y}_n}$. На основании этих равенств и полученных выражений для коэффициентов $\varkappa_{ij}$ уравнения (19) для оптимальных значений α_n и β_n принимают вид

$$\begin{aligned} &\alpha_n K_{\hat{y}_n} + \beta_n b_n K_{\hat{y}_n} = a_n K_{\hat{y}_n}, \\ &a_n K_{\hat{y}_n} b_n^T + \beta_n (b_n K_{y_n} b_n^T + \psi_n' G_n \psi_n'^T) = a_n K_{y_n} b_n^T + \psi_n G_n \psi_n'^T. \end{aligned} \tag{26}$$

Эти уравнения легко решаются. Вычитая первое уравнение (26), умноженное справа на матрицу b_n^T, из второго, получаем

$$\beta_n [b_n (K_{y_n} - K_{\hat{y}_n}) b_n^T + \psi_n' G_n \psi_n'^T] = \\ = a_n (K_{y_n} - K_{\hat{y}_n}) b_n^T + \psi_n G_n \psi_n'^T.$$

Заметим теперь, что в силу равенств $K_{y_n\hat{y}_n} = K_{\hat{y}_n y_n} = K_{\hat{y}_n}$, $M\hat{Y}_n = MY_n$ величина $K_{y_n} - K_{\hat{y}_n}$ представляет собой второй момент ошибки R_n. Действительно,

$$R_n = M(\hat{Y}_n - Y_n)(\hat{Y}_n^T - Y_n^T) = M(\hat{Y}_n^0 - Y_n^0)(\hat{Y}_n^T - Y_n^T) = \\ = K_{\hat{y}_n} - K_{\hat{y}_n y_n} - K_{y_n\hat{y}_n} + K_{y_n} = K_{y_n} - K_{\hat{y}_n}.$$

Следовательно, уравнение для оптимального значения β_n имеет вид

$$\beta_n (b_n R_n b_n^T + \psi_n' G_n \psi_n'^T) = a_n R_n b_n^T + \psi_n G_n \psi_n'^T.$$

Решая это уравнение относительно β_n, находим

$$\beta_n = (a_n R_n b_n^T + \psi_n G_n \psi_n'^T)(b_n R_n b_n^T + \psi_n' G_n \psi_n'^T)^{-1}. \tag{27}$$

Заметим, что матрица $b_n R_n b_n^T$ в задачах практики всегда обратима, так как ее необратимость равноценна существованию линейной функции вектора ошибки $\hat{Y}_n - Y_n$ с нулевой дисперсией. Это означало бы, что соответствующая линейная функция вектора Y_n могла бы быть оценена абсолютно точно, что невозможно. Тем более обратима матрица $b_n R_n b_n^T + \psi_n' G_n \psi_n'^T$. Поэтому формула (27) всегда определяет оптимальное значение β_n.

После определения β_n первое уравнение (26) дает $\alpha_n = a_n - \beta_n b_n$ при условии, что матрица $K_{\hat{y}_n}$ обратима. Это условие тоже всегда выполняется в задачах прак-

тики, так как в случае его невыполнения существовала бы линейная функция вектора Y_n, оценка которой имела бы нулевую дисперсию, что невозможно.

Подставив полученные выражения α_n, m_n, $l_1^{(n)}$ и $l_2^{(n)}$ в (20) и учитывая, что $M\hat{Y}_n = MY_n$, находим оптимальное значение γ_n:

$$\gamma_n = a_n MY_n + a_{n0} - \alpha_n M\hat{Y}_n - \beta_n (b_n MY_n + b_{n0}) = a_{n0} - \beta_n b_{n0}.$$

Наконец, подставив найденные выражения α_n и γ_n в уравнение, определяющее оценки $\hat{Y}_n$, приведем его к виду

$$\hat{Y}_{n+1} = a_n \hat{Y}_n + a_{n0} + \beta_n (X_n - b_n \hat{Y}_n - b_{n0}). \qquad (28)$$

Осталось написать уравнение (22) для характеристической функции $g_n(\lambda, \mu)$. В данном случае оно имеет вид

$$g_n(\lambda, \mu) = h_{n-1}(\psi_{n-1}^T \lambda + \psi_{n-1}'^T \beta_{n-1}^T \mu) \times \\ \times M \exp\{i\lambda^T (a_{n-1} Y_{n-1} + a_{n-1,0}) + \\ + i\mu^T (\alpha_{n-1} \hat{Y}_{n-1} + \beta_{n-1} b_{n-1} Y_{n-1} + a_{n-1,0})\} = \\ = h_{n-1}(\psi_{n-1}^T \lambda + \psi_{n-1}'^T \beta_{n-1}^T \mu) \exp\{i(\lambda^T + \mu^T) a_{n-1,0}\} \times \\ \times M \exp\{i(\lambda^T a_{n-1} + \mu^T \beta_{n-1} b_{n-1}) Y_{n-1} + i\mu^T \alpha_{n-1} \hat{Y}_{n-1}\}$$

или

$$g_n(\lambda, \mu) = h_{n-1}(\psi_{n-1}^T \lambda + \psi_{n-1}'^T \beta_{n-1}^T \mu) \times \\ \times \exp\{i(\lambda^T + \mu^T) a_{n-1,0}\} g_{n-1}(a_{n-1}^T \lambda + b_{n-1}^T \beta_{n-1}^T \mu,\ \alpha_{n-1}^T \mu).$$

Из этой формулы легко выводится рекуррентная формула для характеристической функции $g_n'(\lambda)$ ошибки $\hat{Y}_n - Y_n$. Заметим, что

$$g_n'(\mu) = M \exp\{i\mu^T (\hat{Y}_n - Y_n)\} = g_n(-\mu, \mu),$$

положим в формуле для $g_n(\lambda, \mu)$ $\lambda = -\mu$. Тогда, учитывая, что $-a_{n-1}^T \mu + b_{n-1}^T \beta_{n-1}^T \mu = -\alpha_{n-1}^T \mu$ и что $g_{n-1}(-\alpha_{n-1}^T \mu, \alpha_{n-1}^T \mu) = g_{n-1}'(\alpha_{n-1}^T \mu)$, получим

$$g_n'(\mu) = h_{n-1}((\psi_{n-1}'^T \beta_{n-1}^T - \psi_{n-1}^T) \mu) g_{n-1}'(\alpha_{n-1}^T \mu).$$

Эта формула полностью определяет распределение ошибки $\hat{Y}_n - Y_n$ на каждом шаге и, следовательно, дает возможность находить доверительные области для Y_n.

Из формулы для характеристической функции $g_n'(\mu)$ ошибки легко выводится разностное уравнение для второго момента ошибки R_n. Для этого достаточно дважды

продифференцировать формулу для $g'_n(\mu)$ по μ и положить после этого $\mu=0$. Предоставив выполнить это читателю, мы дадим здесь другой вывод уравнения для R_n.

Вычитая уравнение (23) из уравнения (28) для оценки и заменив X_n его выражением (24), получаем для ошибки $\tilde{Y}_n=\hat{Y}_n-Y_n$ уравнение

$$\tilde{Y}_{n+1}=(a_n-\beta_n b_n)\tilde{Y}_n+(\beta_n\psi'_n-\psi_n)V_n. \qquad (29)$$

Отсюда, принимая во внимание, что Y_n и $\hat{Y}_n$, а следовательно, и $\tilde{Y}_n$, не зависят от V_n, и применяя формулу (3.50) для ковариационной матрицы линейной функции случайного вектора, получаем

$$R_{n+1}=(a_n-\beta_n b_n)R_n(a_n^T-b_n^T\beta_n^T)+$$
$$+(\beta_n\psi'_n-\psi_n)G_n(\psi'^T_n\beta_n^T-\psi_n^T).$$

Раскрыв в каждом слагаемом вторые скобки и пользуясь формулой (27), определяющей β_n, можно привести полученное уравнение к виду

$$R_{n+1}=(a_n-\beta_n b_n)R_n a_n^T+(\psi_n-\beta_n\psi'_n)G_n\psi_n^T. \qquad (30)$$

Это уравнение вместе с формулой (27) для β_n определяет рекуррентный процесс для последовательного вычисления матриц R_n и β_n. Характеристические функции ошибок $g'_n(\mu)$ при этом можно не вычислять. Они могут понадобиться лишь для определения доверительных областей для Y_n.

Определяемая уравнением (29) оценка $\hat{Y}_{n+1}$ при каждом n оптимальна в классе линейных функций (со сдвигом) величин $\hat{Y}_n$ и X_n. Докажем, что она оптимальна и в более широком классе линейных функций величин $\hat{Y}_1, X_1, \ldots, X_n$. Согласно результатам п. 9.2.5 для этого достаточно показать, что для $\hat{Y}_{n+1}$ выполняются соотношения (9.11), которые в данном случае имеют вид

$$M(\hat{Y}_{n+1}-Y_{n+1})\hat{Y}_1^T=0, \quad M(\hat{Y}_{n+1}-Y_{n+1})X_p^T=0$$
$$(p=1, \ldots, n).$$

Подставив в эти формулы выражение $\hat{Y}_{n+1}-Y_{n+1}=\tilde{Y}_{n+1}$ из уравнения для ошибки (29) и приняв во внимание

независимость V_n от $\hat{Y}_1, X_1, \ldots, X_{n-1}$, получаем

$$M(\hat{Y}_{n+1}-Y_{n+1})\hat{Y}_1^T=(a_n-\beta_n b_n)M(\hat{Y}_n-Y_n)\hat{Y}_1^T,$$
$$M(\hat{Y}_{n+1}-Y_{n+1})X_p^T=(a_n-\beta_n b_n)M(\hat{Y}_n-Y_n)X_p^T$$
$$(p=1, \ldots, n-1).$$

Из этих равенств и второго соотношения (25) следует, что оценка $\hat{Y}_{n+1}$ оптимальна в классе линейных функций величин $\hat{Y}_1, X_1, \ldots, X_n$, если оценка $\hat{Y}_n$ оптимальна в классе линейных функций величин $\hat{Y}_1, X_1, \ldots, X_{n-1}$. А так как оценка $\hat{Y}_2=a_1Y_1+a_{10}+\beta_n(X_1-b_1\hat{Y}_1-b_{10})$ оптимальна в классе линейных функций величин $\hat{Y}_1, X_1, \ldots$ то и при любом n оценка $\hat{Y}_{n+1}$ оптимальна в классе линейных функций величин $\hat{Y}_1, X_1, \ldots, X_n$.

Более того, если все величины V_n и величины Y_1 и $\hat{Y}_1$ распределены нормально, то и величины $Y_{n+1}, \hat{Y}_1, X_1, \ldots$ $\ldots, X_{n+1}$ при любом n имеют совместное нормальное распределение. В этом случае оценка $\hat{Y}_{n+1}$ совпадает с регрессией Y_{n+1} на $\hat{Y}_1, X_1, \ldots, X_{n+1}$ и, следовательно, оптимальна в классе всех функций величин $\hat{Y}_1, X_1, \ldots$ $\ldots, X_{n+1}$ (абсолютно оптимальна).

2.4. Фильтры Калмана. Изложенный метод линейного оценивания случайных величин, определяемых линейной моделью авторегрессии (23), разработан Калманом [33]. Поэтому автоматические системы, реализующие вычисление оценок, определяемых уравнением (28), обычно называются фильтрами Калмана.

Рис. 35.

Построение фильтра Калмана по данной модели формирования последовательности $\{X_n\}$ можно представить наглядной схемой, если воспользоваться обычным в теории управления способом изображения систем и их соединений, указывая линиями со стрелками направление передачи сигналов. На рис. 35 прямоугольником с буквами MA показана модель авторегрессии (23) со стрелками, указывающими входной сигнал V_n и выходной сигнал Y_{n+1}. Из сравнения (28) с (23) видно, что та же модель (23) будет давать на выходе последовательность оценок $\{\hat{Y}_n\}$, если вместо $\{V_n\}$ подать на ее вход после-

довательность $\{\beta_n(X_n - b_n\hat{Y}_n - b_{n0})\}$. Чтобы осуществить это, следует выходной сигнал $\hat{Y}_{n+1}$ подвести к входу по цепи обратной связи, содержащей устройство, осуществляющее умножение $\hat{Y}_n$ на b_n*). Перед входом в модель (23) величина $b_n\hat{Y}_n$ должна вычитаться из соответствующей величины $X_n - b_{n0}$ и разность должна умножаться на соответствующий коэффициент β_n. Результат этого построения показан на рис. 36. Кружком, разделенным на

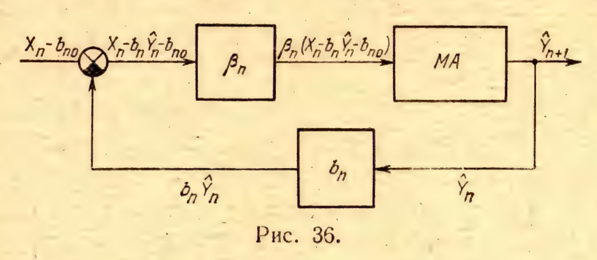


Рис. 36.

квадранты, показан сумматор. Зачерненный нижний квадрант перед суммированием изменяет знак поступающего в него сигнала (отрицательная обратная связь). Прямоугольниками внизу и после сумматора показаны устройства, выполняющие умножение на матрицы b_n и β_n соответственно (усилители с коэффициентами усиления b_n и β_n, в общем случае матричными). Входным сигналом построенной таким путем системы служит последовательность случайных величин $\{X_n - b_{n0}\}$.

Таким образом, фильтр Калмана получается из исходной модели авторегрессии (23) замыканием ее отрицательной обратной связью, содержащей усилитель с коэффициентом усиления b_n, установкой после сумматора другого усилителя с коэффициентом усиления β_n и подачей на вход полученной таким путем системы последовательности случайных величин $\{X_n - b_{n0}\}$ (рис. 36). Коэффициенты усиления b_n и β_n, очевидно, в общем случае будут переменными, т. е. будут изменяться на каждом шаге работы модели.

---

*) Предполагается, что время прохождения сигнала по цепи обратной связи равно длительности одного шага (такта), так что к моменту появления на выходе величины $\hat{Y}_{n+1}$ по цепи обратной связи идет предыдущая величина $\hat{Y}_n$.

Пример 2. Последовательность величин $\{Y_n\}$ определяется моделью авторегрессии

$$Y_{n+2}+a_{1n}Y_{n+1}+a_{2n}Y_n=b_{1n}V_{n+1}+b_{0n}V_n.$$

Построить фильтр Калмана для оценок величин Y_n по результатам их наблюдения с ошибками, образующими последовательность независимых случайных величин $\{U_n\}$.

В данном случае $X_n=Y_n+U_n$. Для решения задачи необходимо сначала привести модель, формирующую последовательность $\{Y_n\}$, к модели авторегрессии первого порядка. По формулам п. 1.3 находим коэффициенты q_{1n}, q_{2n} и q_{3n}:

$$q_{1n}=b_{2,\,n-2}=0, \qquad q_{2n}=b_{1,\,n-1}+a_{1,\,n-1}q_{1n}=b_{1,\,n-1},$$
$$q_{3n}=b_{0n}+a_{0n}q_{1n}+a_{1n}q_{2n}=b_{0n}+a_{1n}b_{1,\,n-1}.$$

После этого находим

$$Z_{1n}=Y_n, \qquad Z_{1,\,n+1}=Z_{2n}+b_{1,\,n-1}V_n.$$

Вводя векторы

$$Z_n=\begin{bmatrix} Y_n \\ Y_{n+1}-b_{1,\,n-1}V_n \end{bmatrix}, \qquad W_n=\begin{bmatrix} b_{1,\,n-1} \\ b_{0n}+b_{1,\,n-1} \end{bmatrix} V_n,$$

получим разностное уравнение первого порядка

$$Z_{n+1}=\begin{bmatrix} 0 & I \\ a_{0n} & a_{1n} \end{bmatrix} Z_n+W_n.$$

Уравнение, определяющее наблюдаемые величины, принимает вид

$$X_n=[I \;\; 0]\, Z_n+U_n.$$

Ковариационная матрица K_n вектора W_n выражается через ковариационную матрицу H_n вектора V_n формулой (3.50):

$$K_n=\begin{bmatrix} b_{1,\,n-1} \\ b_{0n}+b_{1,\,n-1} \end{bmatrix} H_n\,[b^T_{1,\,n-1} \;\; b^T_{0n}+b^T_{1,\,n-1}].$$

Так как $W_n=[I \;\; 0]\,[W_n^T \;\; U_n^T]^T$, $U_n=[0 \;\; I]\,[W_n^T \;\; U_n^T]^T$, то в данном случае $\psi_n=[I \;\; 0]$, $\psi'_n=[0 \;\; I]$. Ковариационная матрица вектора $[W_n^T \;\; U_n^T]^T$, играющая роль матрицы G_n в общей теории, имеет в данном случае вид

$$G'_n=\begin{bmatrix} K_n & 0 \\ 0 & G_n \end{bmatrix}.$$

Уравнение (28) в данном случае имеет вид

$$\hat{Z}_{n+1}=\left(\begin{bmatrix} 0 & I \\ a_{0n} & a_{1n} \end{bmatrix}-\beta_n\,[I \;\; 0]\right)\hat{Z}_n+\beta_n X_n,$$

где на основании (27)

$$\beta_n=\begin{bmatrix} 0 & I \\ a_{0n} & a_{1n} \end{bmatrix} R_n \begin{bmatrix} I \\ 0 \end{bmatrix}\left([I \;\; 0]\, R_n \begin{bmatrix} I \\ 0 \end{bmatrix}+G_n\right)^{-1},$$

а $\hat{R}_n$ определяется уравнением (30), которое в данном случае имеет вид

$$R_{n+1}=\left(\begin{bmatrix}0 & I\\ a_{0n} & a_{1n}\end{bmatrix}-\beta_n\,[I\ \ 0]\right)R_n\begin{bmatrix}0 & a_{0n}\\ I & a_{1n}\end{bmatrix}+K_n.$$

Обратим внимание на то, что в данном случае фильтр Калмана дает не только оценки интересующих нас случайных величин Y_n, но еще и оценки величин $Z_{2n}=Y_{n+1}-b_{1,\,n-1}V_n$. Это будет всегда, когда исходная модель авторегрессии выше первого порядка. Однако в общем случае эти оценки бесполезны из-за наличия в них величин V_n. Поэтому их обычно не используют (не выводят на выходные устройства ЭВМ). И лишь в частных случаях эти дополнительные оценки могут быть полезными. Так, например, если $b_{1n}=0$ при всех n, то $Z_{2n}=Y_{n+1}$ и, следовательно, фильтр Калмана дает, кроме оценки $\hat{Y}_n$, и оптимальный прогноз на шаг вперед величины Y_{n+1} (заметим, что при этом $\hat{Z}_{2n}\neq\hat{Y}_{n+1}$).

2.5. Обновляющие последовательности. Обратим внимание на то, что модель (28) фактически использует для формирования оценок $\hat{Y}_n$ не саму последовательность наблюдаемых величин $\{X_n\}$, а последовательность некоррелированных случайных величин $\{X_n-b_n\hat{Y}_n-b_{n0}\}$ *). Эта последовательность получается в результате вычитания из каждой величины X_n ее оптимальной несмещенной линейной оценки $\hat{X}_n$, полученной по результатам наблюдения предшествующих величин $X_1, \ldots, X_{n-1}$. При этом от каждой величины X_n остается только та часть, которая не может быть линейно выражена через $X_1, \ldots, X_{n-1}$ и поэтому содержит новую информацию по сравнению с полученной ранее в результате наблюдения $X_1, \ldots, X_{n-1}$.

Последовательность некоррелированных случайных величин, полученная из данной последовательности путем вычитания из каждой величины ее оптимальной линейной оценки по результатам наблюдения предшествующих величин, называется обновляющей последовательностью. Пользуясь этим понятием, можно сформулировать полученный результат совсем просто: последовательность оптимальных оценок $\{\hat{Y}_n\}$ определяется той же линейной регрессионной моделью, что и последова-

*) Из (24) следует, что $W_n=X_n-b_n\hat{Y}_n-b_{n0}=b_n(Y_n-\hat{Y}_n)+\psi_n' V_n$. Отсюда, принимая во внимание, что $Y_n-\hat{Y}_n$ и V_n не коррелированы с $\hat{Y}_1, X_1, \ldots, X_{n-1}$, заключаем, что каждая величина W_n не коррелирована с $\hat{Y}_1, X_1, \ldots, X_{n-1}$, а следовательно, и с $\hat{Y}_p$ при $p\leqslant n$, поскольку $\hat{Y}_p$ является линейной функцией $\hat{Y}_1, X_1, \ldots, X_{p-1}$. Следовательно, каждая величина W_n не коррелирована с W_p при $p<n$.

тельность $\{Y_n\}$, с заменой последовательности $\{V_n\}$ обновляющей последовательностью $\{X_n - b_n\hat{Y}_n - b_{n0}\}$, каждый член которой должен быть умножен на соответствующий коэффициент β_n.

2.6. Приближенное решение нелинейных задач оценивания. Метод п. 2.1 и п. 2.2 дает точное решение задачи нахождения условно оптимальных оценок $\hat{Y}_n$ векторов Y_n, минимизирующих средний квадрат ошибки $M|\hat{Y}_n - Y_n|^2$ при условии, что они удовлетворяют разностному уравнению данного вида. Однако полученное точное решение задач нелинейного оценивания требует весьма громоздких вычислений, связанных с определением на каждом шаге характеристической функции g_n. Поэтому большое практическое значение имеет разработка приближенных методов вычисления характеристических функций $g_n(\lambda, \mu)$ и соответствующих плотностей

$$f_n(y, \hat{y}) = \frac{1}{(2\pi)^{2m}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i\lambda^T y - i\mu^T \hat{y}} g_n(\lambda, \mu)\, d\lambda\, d\mu$$

(m — размерность векторов Y_n и $\hat{Y}_n$).

Для приближенного определения характеристических функций g_n и соответствующих плотностей f_n можно предложить следующий прием: заменить неизвестную функцию $g_n(\lambda, \mu)$ некоторой вполне определенной функцией $g(\lambda, \mu, \sigma_n)$, зависящей от конечного числа неизвестных параметров, образующих вектор σ_n, и вывести из уравнения (12) соответствующие приближенные рекуррентные формулы (разностные уравнения), определяющие последовательность значений неизвестных параметров $\{\sigma_n\}$ *).

Заменив $g_n(\lambda, \mu)$ функцией $g(\lambda, \mu, \sigma_n)$ и плотность $f_{n-1}(y, \hat{y})$ соответствующей плотностью $f(y, \hat{y}, \sigma_{n-1})$, заменим формулу (12) приближенной формулой

$$g(\lambda, \mu, \sigma_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{i\lambda^T \omega_{n-1}(y, v) + \\ + i\mu^T [\delta_{n-1}\zeta_{n-1}(\hat{y}, \omega'_{n-1}(y, v)) + \gamma_{n-1}]\} \times \\ \times p_{n-1}(v) f(y, \hat{y}, \sigma_{n-1})\, dv\, dy\, dy_{n-1}.$$

---

*) Эта идея принадлежит В. И. Демуху, который положил ее в основу разработанного им приближенного метода анализа точности нелинейных систем («Автоматика и телемеханика» 1965, № 6, с. 1021—1025).

Отсюда дифференцированием по λ и μ и приравниванием λ и μ нулю можно получить приближенные формулы для моментов случайного вектора $[Y_n^T \ \hat{Y}_n^T]^T$:

$$\alpha^{(n)}_{k_1, \ldots, k_m, l_1, \ldots, l_m} = \left[\frac{\partial^{k_1 + \ldots + k_m + l_1 + \ldots + l_m}}{\partial (i\lambda_1)^{k_1} \ldots \partial (i\mu_m)^{l_m}} g(\lambda, \mu, \sigma_n) \right]_0,$$

где индексом 0 внизу отмечено значение величины в квадратных скобках при $\lambda = \mu = 0$. Учитывая, что эти моменты зависят от σ_n, напишем полученные приближенные формулы для моментов в виде

$$\alpha^{(n)}_{k_1, \ldots, k_m, l_1, \ldots, l_m}(\sigma_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial^{k_1 + \ldots + k_m + l_1 + \ldots + l_m}}{\partial (i\lambda_1)^{k_1} \ldots \partial (i\mu_m)^{l_m}} \exp \{ i\lambda^T \omega_{n-1}(y, v) + i\mu^T [\delta_{n-1} \zeta_{n-1}(\hat{y}, \omega'_{n-1}(y, v)) + \gamma_{n-1}] \} \right]_0 \times p_{n-1}(v) f(y, \hat{y}, \sigma_{n-1}) \, dv \, dy \, d\hat{y}. \quad (31)$$

Взяв столько таких уравнений, сколько координат имеют векторы σ_n и σ_{n-1}, получим систему разностных уравнений для приближенного определения последовательности векторов $\{\sigma_n\}$.

В частном случае, когда σ_n представляет собой совокупность моментов вектора $[Y_n^T \ \hat{Y}_n^T]^T$ до порядка r включительно, $\sigma_n = \alpha^{(n)}$, уравнения (31) дают рекуррентные формулы для приближенного вычисления моментов:

$$\alpha^{(n)}_{k_1, \ldots, k_m, l_1, \ldots, l_m} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial^{k_1 + \ldots + l_m}}{\partial (i\lambda_1)^{k_1} \ldots \partial (i\mu_m)^{l_m}} \times \exp \{ i\lambda^T \omega_{n-1}(y, v) + i\mu^T [\delta_{n-1} \zeta_{n-1}(\hat{y}, \omega_{n-1}(y, v)) + \gamma_{n-1}] \} \right]_0 p_{n-1}(v) f(y, \hat{y}, \alpha^{(n-1)}) \, dv \, dy \, d\hat{y}$$

$$(k_1, \ldots, k_m, l_1, \ldots, l_m = 0, 1, \ldots, r; \quad k_1 + \ldots + k_m + l_1 + \ldots + l_m \leqslant r).$$

Пример 3. Заменив в уравнении для $g_n(\lambda, \mu)$ примера 1 плотность $f_{n-1}(y, \eta)$ нормальной плотностью, определяемой первыми и вторыми моментами $\alpha_{10}^{(n-1)}$, $\alpha_{01}^{(n-1)}$, $\alpha_{20}^{(n-1)}$, $\alpha_{11}^{(n-1)}$, $\alpha_{02}^{(n-1)}$, получим изложенным методом пять рекуррентных формул, определяющих приближенно моменты $\alpha_{10}^{(n)}$, $\alpha_{01}^{(n)}$, $\alpha_{20}^{(n)}$, $\alpha_{11}^{(n)}$, $\alpha_{02}^{(n)}$ по известным $\alpha_{10}^{(n-1)}$, $\alpha_{01}^{(n-1)}$, $\alpha_{20}^{(n-1)}$, $\alpha_{11}^{(n-1)}$, $\alpha_{02}^{(n-1)}$. Эти формулы можно вывести значи-

тельно проще, взяв все первые и вторые производные по λ и μ выражения для $g_n(\lambda, \mu)$, положив после этого $\lambda=\mu=0$ и выразив в полученных формулах все моменты выше второго порядка через первые и вторые моменты по формулам для нормального распределения (п. 4.5.3). Все моменты в формулах для оптимальных значений δ_n и γ_n выразятся при этом через $\alpha_{10}^{(n)}$, $\alpha_{01}^{(n)}$, $\alpha_{20}^{(n)}$, $\alpha_{11}^{(n)}$, $\alpha_{02}^{(n)}$, так что необходимость определять эти моменты по формулам примера 1 отпадает. Момент второго порядка ошибки $\hat{Y}_n - Y_n$ в этом случае определяется формулой

$$R_n = \alpha_{20}^{(n)} - \alpha_{11}^{(n)}.$$

2.7. Оценивание неизвестных параметров в разностных уравнениях. Метод п. 2.1 и п. 2.2 дает возможность оценивать случайные величины, определяемые разностным уравнением, и в том случае, когда функции ω_n и ω_n' в (8) и (9) зависят от конечного числа неизвестных параметров (конечномерного векторного параметра). При этом оказывается возможным попутно оценивать и неизвестные параметры.

Предположим, что функции ω_n и ω_n' в (8) и (9) (функции φ_n, ψ_n, φ_n' и ψ_n' в (16) и (17) в частном случае п. 2.2) зависят от неизвестного конечномерного векторного параметра θ. В этом случае уравнение (8) и формулу (9) можно написать в виде

$$Y_{n+1} = \omega_n(Y_n, \theta, V_n), \qquad X_n = \omega_n'(Y_n, \theta, V_n).$$

Поставим задачу найти оптимальные оценки $\hat{Y}_n$ и $\hat{\Theta}_n$ вектора Y_n и параметра θ на каждом шаге, определяемые разностными уравнениями вида (10) по результатам наблюдения последовательности векторов $\{X_n\}$, т. е. задачу одновременного оценивания векторов Y_n и неизвестного параметра θ.

Подобные задачи обычно решаются сведением их к задаче п. 2.1 с полностью известными функциями ω_n и ω_n'. Для этого заменяют неизвестный параметр θ последовательностью случайных векторов $\{\Theta_n\}$, $\Theta_n = \theta$ (распределение каждого из этих векторов сосредоточено в одной точке θ). Очевидно, что введенная таким путем последовательность векторов $\{\Theta_n\}$ удовлетворяет разностному уравнению $\Theta_{n+1} = \Theta_n$. В результате уравнение формирования последовательности векторов $\{Y_n\}$ заменится системой разностных уравнений

$$Y_{n+1} = \omega_n(Y_n, \Theta_n, V_n), \qquad \Theta_{n+1} = \Theta_n,$$

а формула, определяющая последовательность наблюдаемых величин $\{X_n\}$, заменится формулой

$$X_n = \omega_n' (Y_n, \Theta_n, V_n).$$

Полученные уравнения формирования последовательностей векторов $\{Y_n\}$ и $\{\Theta_n\}$ можно записать в виде одного уравнения типа (8) для последовательности составных векторов $\{Z_n\}$, $Z_n = [Y_n^T \ \Theta_n^T]^T$, с полностью известными функциями ω_n, а формула для X_n представляет собой формулу (9) с заменой вектора Y_n составным вектором $Z_n = [Y_n^T \ \Theta_n^T]^T$ с полностью известной функцией ω_n'.

Таким образом, задача одновременного оценивания величин, определяемых разностным уравнением, и неизвестных параметров в этом уравнении сводится к задаче, решенной в п. 2.1, путем введения составных векторов, каждый из которых включает все координаты соответствующего вектора Y_n, определяемого разностным уравнением, и все неизвестные параметры в этом уравнении.

Естественно, разностное уравнение для составных векторов оценок при этом будет иметь вид

$$[\hat{Y}_{n+1}^T \ \hat{\Theta}_{n+1}^T]^T = \delta_n \zeta_n (\hat{Y}_n, \hat{\Theta}_n, X_n) + \gamma_n.$$

Аналогично запишется уравнение для оценок в частном случае п. 2.2:

$$[\hat{Y}_{n+1}^T \ \hat{\Theta}_{n+1}^T]^T = \alpha_n \xi_n (\hat{Y}_n, \hat{\Theta}_n) + \beta_n \eta_n (\hat{Y}_n, \hat{\Theta}_n) X_n + \gamma_n.$$

Если требуется оценивать не все координаты векторов Y_n, а только часть их, то в уравнении для оценок следует взять функции ζ_n (ξ_n и η_n в частном случае п. 2.2), зависящие только от соответствующей части координат вектора $\hat{Y}_n$ (очевидно, что при этом необходимо оценивать все неизвестные параметры). В частности, при желании оценивать только неизвестные параметры в разностном уравнении функции ζ_n (ξ_n и η_n в частном случае п. 2.2) следует взять не зависящими от $\hat{Y}_n$.

Пример 4. Последовательность скалярных случайных величин $\{Y_n\}$ определяется уравнением

$$Y_{n+1} = \theta Y_n + V_n,$$

где θ — неизвестный параметр, $\{V_n\}$ — последовательность независимых случайных величин. Найти оптимальные линейные оценки $\hat{\Theta}_n$, $\hat{Y}_n$ параметра θ и величин Y_n, определяемые уравнением

$$[\hat{Y}_{n+1} \ \hat{\Theta}_{n+1}]^T = \alpha_n [\hat{Y}_n \ \hat{\Theta}_n]^T + \beta_n X_n + \gamma_n,$$

по результатам наблюдения величин $X_n = Y_n + U_n$, где $\{U_n\}$ — последовательность независимых случайных величин, независимая от $\{V_n\}$.

Оптимальные значения α_n, β_n и γ_n в этом случае определяются системой линейных алгебраических уравнений (19) при

$$\varkappa_{01}^{(n)} = \begin{bmatrix} MY_n\Theta_n\hat{Y}_n - MY_n\Theta_n M\hat{Y}_n & MY_n\Theta_n\hat{\Theta}_n - MY_n\Theta_n M\hat{\Theta}_n \\ M\Theta_n\hat{Y}_n - M\Theta_n M\hat{Y}_n & M\Theta_n\hat{\Theta}_n - M\Theta_n M\hat{\Theta}_n \end{bmatrix},$$

$$\varkappa_{02}^{(n)} = [MY_n^2\Theta_n - MY_n\Theta_n MY_n \quad MY_n\Theta_n - MY_n M\Theta_n]^T,$$

$$\varkappa_{11}^{(n)} = \begin{bmatrix} M\hat{Y}_n^2 - (M\hat{Y}_n)^2 & M\hat{Y}_n\hat{\Theta}_n - M\hat{Y}_n M\hat{\Theta}_n \\ M\hat{Y}_n\hat{\Theta}_n - M\hat{Y}_n M\hat{\Theta}_n & M\hat{\Theta}_n^2 - (M\hat{\Theta}_n)^2 \end{bmatrix},$$

$$\varkappa_{12}^{(n)} = \varkappa_{21}^{(n)T} = [MY_n\hat{Y}_n - MY_n M\hat{Y}_n \quad MY_n\hat{\Theta}_n - MY_n M\hat{\Theta}_n]^T,$$

$$\varkappa_{22}^{(n)} = MY_n^2 - (MY_n)^2 + G_n,$$

где G_n — дисперсия величины U_n. Оптимальное значение γ_n определяется формулой (20):

$$\gamma_n = [MY_n\Theta_n \quad M\Theta_n]^T - \alpha_n [M\hat{Y}_n \quad M\hat{\Theta}_n]^T - \beta_n MY_n.$$

Уравнение (22) для характеристической функции величин Y_n, Θ_n, $\hat{Y}_n$, $\hat{\Theta}_n$ имеет вид

$$g_n(\lambda, \mu) = h_{n-1}(\lambda_1)\, l_{n-1}(\beta_{n-1}^T\mu) \times$$
$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{i\lambda_1 y\theta + i\lambda_2\theta + i\mu^T(\alpha_{n-1}[y' \ \theta']^T +$$
$$+ \beta_{n-1}y + \gamma_{n-1})\}\, f_{n-1}(y, \theta, y', \theta')\, dy\, d\theta\, dy'\, d\theta',$$

где $\lambda = [\lambda_1 \ \lambda_2]^T$, $\mu = [\mu_1 \ \mu_2]^T$, $h_{n-1}(\lambda)$ и $l_{n-1}(\lambda')$ — характеристические функции величин V_{n-1} и U_{n-1} соответственно, а $f_{n-1}(y, \theta, y', \theta')$ — плотность случайного вектора $[Y_{n-1}\Theta_{n-1} \ \hat{Y}_{n-1} \ \hat{\Theta}_{n-1}]^T$:

$$f_{n-1}(y, \theta, y', \theta') = \frac{1}{16\pi^4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(\lambda_1 y + \lambda_2\theta + \mu_1 y' + \mu_2\theta')} \times$$
$$\times g_{n-1}(\lambda_1, \lambda_2, \mu_1, \mu_2)\, d\lambda_1\, d\lambda_2\, d\mu_1\, d\mu_2.$$

Моменты величин Y_n, Θ_n, $\hat{Y}_n$, $\hat{\Theta}_n$, входящие в полученные формулы для $\varkappa_{ij}^{(n)}$ и γ_n, можно вычислить, как и в примере 1, или дифференцированием характеристической функции g_n, или интегрированием, определив предварительно плотность $f_n(y, \theta, y', \theta')$.

Для приближенного решения задачи методом п. 2.6 можно продифференцировать уравнение для $g_n(\lambda_1, \lambda_2, \mu_1, \mu_2)$ по λ_1, λ_2, μ_1, μ_2 и по всем парам этих переменных, положить после этого $\lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 0$ и заменить все моменты выше второго порядка их выражениями через моменты первого и второго порядков по формулам п. 4.5.3 для нормального распределения. Тогда получим рекуррентные формулы для определения моментов первого и второго порядков вектора $[Y_n\Theta_n \ \hat{Y}_n\hat{\Theta}_n]^T$. Через эти моменты выразятся и все моменты в формулах для $\varkappa_{ij}^{(n)}$ и γ_n, а также матрица R_n вторых моментов ошибки $\hat{Y}_n - Y_n$.

Пример 5. В условиях примера 4 найти линейную оценку одного параметра θ, определяемую уравнением

$$\hat{\Theta}_{n+1}=\alpha_n\hat{\Theta}_n+\beta_n X_n+\gamma_n.$$

В этом случае

$$\varkappa_{01}^{(n)}=[MY_n\Theta_n\hat{\Theta}_n-MY_n\Theta_n M\hat{\Theta}_n \;\; M\Theta_n\hat{\Theta}_n-M\Theta_n M\hat{\Theta}_n]^T,$$
$$\varkappa_{12}^{(n)}=\varkappa_{21}^{(n)T}=MY_n\hat{\Theta}_n-MY_n M\hat{\Theta}_n, \qquad \varkappa_{11}^{(n)}=M\hat{\Theta}_n-(M\hat{\Theta}_n)^2,$$

а $\varkappa_{02}^{(n)}$ и $\varkappa_{22}^{(n)}$ имеют те же значения, что и в предыдущем примере. Оптимальное значение γ_n определяется формулой

$$\gamma_n=M\Theta_n-\alpha_n M\hat{\Theta}_n-\beta_n MY_n.$$

Уравнение (22) для характеристической функции величин Y_n, Θ_n, $\hat{\Theta}_n$ имеет вид

$$g_n(\lambda,\ \mu)=h_{n-1}(\lambda_1)\, l_{n-1}(\beta_{n-1}\mu)\times$$
$$\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\{i\lambda_1 y\theta+i\lambda_2\theta+i\mu(\alpha_{n-1}\theta'+$$
$$+\beta_{n-1}y+\gamma_{n-1})\}\, f_{n-1}(y,\ \theta,\ \theta')\, dy\, d\theta\, d\theta',$$

где

$$f_{n-1}(y,\theta,\theta')=\frac{1}{8\pi^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\{-i(\lambda_1 y+\lambda_2\theta+\mu\theta')\}\times$$
$$\times g_{n-1}(\lambda_1,\ \lambda_2,\ \mu)\, d\lambda_1\, d\lambda_2\, d\mu.$$

Все сказанное в конце примера 4 относительно приближенного определения $g_n(\lambda,\ \mu)$ полностью относится и к данному случаю.

Метод п. 2.1 и п. 2.2 позволяет не только оценивать неизвестные параметры в системах, описываемых известными разностными уравнениями, но и строить статистические модели систем, не поддающихся математическому описанию с помощью каких-либо уравнений.

Отобрав существенные переменные, определяющие состояние системы в каждый данный момент, и связав их разностным уравнением подходящей формы, содержащим достаточное число неизвестных параметров, можно по результатам наблюдения работы системы (производимых в ней измерений) оценить изложенным в п. 2.1 и п. 2.2 методом все неизвестные параметры и таким путем получить статистическую модель системы, описываемую разностным уравнением. Взяв различные разностные уравнения, можно параллельно оценивать неизвестные параметры в них, сравнивать соответствующие модели по средним квадратам ошибок оценок параметров и выбрать наиболее подходящую из них.

Таким образом, изложенный здесь метод оценивания величин, определяемых разностными уравнениями, и неизвестных параметров в них имеет широкое поле приложений.

§ 3. Факторные модели

3.1. Задачи факторного анализа. Во многих задачах практики приходится изучать случайные векторы большой размерности. При этом часто оказывается, что распределение случайного вектора сосредоточено вблизи некоторого подпространства значительно меньшей размерности (эллипсоид рассеивания сплющен по некоторым направлениям настолько, что его размерами в этих направлениях можно пренебречь). При этих условиях случайный вектор можно с достаточной для практики точностью представить как результат линейного преобразования случайного вектора меньшей размерности. В связи с этим важной задачей статистики является задача нахождения приближенного линейного представления случайного вектора через вектор меньшей размерности по результатам опытов. Характерная особенность этой задачи состоит в том, что вектор меньшей размерности, через который надо выразить наблюдаемый случайный вектор, не наблюдается и о нем ничего не известно.

Ясно, что решение вопроса о возможности точного представления случайного вектора через вектор меньшей размерности по экспериментальным данным принципиально невозможно. Можно лишь пытаться построить соответствующую модель и определить, согласуется ли гипотеза о том, что распределение сосредоточено на некотором подпространстве, с экспериментальными данными.

Предположим, что в результате n опытов наблюдается m-мерный случайный вектор Y, представляющий собой сумму интересующего нас случайного вектора X и ошибки измерения Z. Допустив, что вектор X можно представить как линейную функцию r-мерного вектора $V = [V_1 \ldots V_r]^T$, $r < m$, можем написать

$$Y = X + Z = m_x + \sum_{p=1}^{r} V_p x_p + Z. \qquad (32)$$

Таким образом, мы приходим к задаче построения модели (32) по результатам наблюдения вектора Y. Методы

решения этой задачи составляют содержание *факторного анализа* [50, 118].

Так как любой случайный вектор можно привести к вектору с некоррелированными координатами, то без потери общности можно считать величины $V_1, \ldots, V_r$ в (32) некоррелированными.

Ошибки измерений координат случайного вектора обычно можно считать некоррелированными как между собой, так и с измеряемым вектором. В этом случае вектор Z не коррелирован с величинами $V_1, \ldots, V_r$ и его ковариационная матрица диагональна.

Ясно, что решение задачи представления наблюдаемого вектора Y формулой (32) неоднозначно. Любое линейное преобразование вектора V, дающее в результате вектор той же размерности r, оставляет структуру модели (32) неизменной. В частности, путем умножения величин $V_1, \ldots, V_r$ на соответствующие числа и векторов $x_1, \ldots, x_r$ на обратные числа можно сделать дисперсии величин $V_1, \ldots, V_r$ равными наперед заданным числам, в частности, все их сделать равными 1.

Таким образом, будем предполагать, что выполнены следующие условия:

1) случайные величины $V_1, \ldots, V_r$ в (32) не коррелированы;

2) ошибка измерений Z имеет нулевое математическое ожидание, не коррелирована с величинами $V_1, \ldots, V_r$, и ее ковариационная матрица диагональна;

3) дисперсии величин $V_1, \ldots, V_r$ равны 1.

Модель наблюдаемого случайного вектора Y, определяемая формулой (32) при выполнении условий 1)—3), называется *факторной моделью*. Случайные величины $V_1, \ldots, V_r$ называются *факторами* *). Координаты вектора x_p называются *нагрузками* соответствующих координат вектора Y (или вектора X) *на p-й фактор* ($p = 1, \ldots, r$). Дисперсии координат вектора X, определяемого суммой в (32), называются *факторными дисперсиями*.

Задача построения факторной модели (32), по существу, равноценна задаче нахождения канонического разложения вектора X (§ 3.4) по результатам наблюдения вектора $Y = X + Z$ (т. е. по результатам наблюдения

*) Термин *фактор* в факторном анализе и в дисперсионном анализе имеет разный смысл.

вектора X с ошибками). Однако непосредственно применить методы § 3.4 для решения этой задачи нельзя, так как для этого необходимо знать ковариационную матрицу вектора X. В данном же случае можно получить по результатам опытов только оценку ковариационной матрицы вектора Y. Чтобы найти подходы к решению поставленной задачи, представим ковариационную матрицу K_x вектора X каноническим разложением (3.75). Тогда ковариационная матрица вектора Y на основании (32) и условий 1)—3) выразится формулой

$$K_y = K_x + K_z = \sum_{p=1}^{r} x_p x_p^T + K_z. \tag{33}$$

В силу диагональности матрицы K_z все недиагональные элементы матрицы K_y совпадают с соответствующими элементами матрицы K_x. Поэтому найденные по результатам опытов оценки недиагональных элементов матрицы K_y можно принять за оценки соответствующих элементов матрицы K_x. Диагональные же элементы матрицы K_x — факторные дисперсии — остаются неизвестными.

Таким образом, задача построения факторной модели (32) сводится к нахождению канонического разложения вектора X при неизвестных диагональных элементах его ковариационной матрицы K_x.

Чтобы преодолеть возникшее затруднение, можно предложить два подхода. Первый подход состоит в том, что вместо канонического разложения вектора X ищут каноническое разложение вектора Y. Определив первые r членов этого разложения, все оставшиеся $m-r$ членов объединяют в ошибку Z, пренебрегая коррелированностью координат полученного таким образом вектора Z. Второй прием состоит в замене неизвестных факторных дисперсий произвольными положительными числами и в сочетании метода канонических разложений с методом итераций.

3.2. Метод главных компонент. На первом подходе основан *метод главных компонент*, использующий разложение наблюдаемого случайного вектора по собственным векторам. В примере 7.13 мы видели, что оценками максимума правдоподобия собственных значений и собственных векторов служат собственные значения и собственные векторы оценки максимума правдоподобия ковариационной матрицы. Если найденные оценки собствен-

ных значений расположить в порядке невозрастания и при этом окажется, что сумма $m-r$ последних собственных значений достаточно мала по сравнению с суммой всех их, то соответствующие главные компоненты можно отнести за счет ошибок измерений, т. е. включить их в вектор Z. При этом разложение (3.66) случайного вектора Y по собственным векторам дает искомую факторную модель (32), если положить $V_p = U_p / \sqrt{\lambda_p}$, $x_p = \varphi_p \sqrt{\lambda_p}$ $(p = 1, \ldots, r)$.

Пример 6. По экспериментальным данным получена оценка ковариационной матрицы K_y, приведенная в примере 3.14. Этой оценке соответствуют найденные в примере 3.14 собственные значения: $\lambda_1 = 6{,}0500$, $\lambda_2 = 0{,}9002$, $\lambda_3 = 0{,}7134$, $\lambda_4 = 0{,}2603$, $\lambda_5 = 0{,}0560$. Эти собственные значения можно принять за оценки соответствующих собственных значений ковариационной матрицы случайного вектора Y.

Так как сумма двух последних собственных значений $\lambda_4 + \lambda_5$ составляет меньше 4% от суммы всех собственных значений, то можно сохранить в разложении только первые три главные компоненты. В результате получим трехфакторную модель

$$\begin{aligned}
Y_1 &= 1{,}354V_1 - 0{,}086V_2 - 0{,}013V_3 + Z_1,\\
Y_2 &= 1{,}310V_1 - 0{,}234V_2 - 0{,}304V_3 + Z_2,\\
Y_3 &= 1{,}325V_1 - 0{,}035V_2 - 0{,}147V_3 + Z_3,\\
Y_4 &= 0{,}840V_1 + 0{,}009V_2 + 0{,}764V_3 + Z_4,\\
Y_5 &= 0{,}502V_1 + 0{,}913V_2 - 0{,}093V_3 + Z_5.
\end{aligned}$$

Метод главных компонент, как и всякий метод, основанный на определении собственных значений и собственных векторов, требует сложных вычислений и поэтому может практически применяться только на ЭВМ.

Другие методы факторного анализа основаны на других канонических разложениях. Все они различаются, по существу, только выбором векторов $g_1, \ldots, g_r$ в общем алгоритме п. 3.4.3.

3.3. Центроидный метод. Одним из часто применяемых методов факторного анализа является центроидный метод [50, 118]. Этод метод основан на следующем выборе каждого очередного вектора g_s. Сначала берут вектор q_s, линейно независимый от предыдущих векторов $q_1, \ldots, q_{s-1}$, все координаты которого равны $+1$ или -1 и определяются так, чтобы число отрицательных слагаемых в квадратичной форме $q_s^T K_x^{(s)} q_s$, где на основании (3.73) и условия 3)

$$K_x^{(1)} = K_x, \quad K_x^{(s)} = K_x - \sum_{p=1}^{s-1} x_p x_p^T \qquad (s = 2, \ldots, r), \quad (34)$$

было минимальным. Затем определяют вектор g_s формулой

$$g_s = \frac{q_s}{\sqrt{q_s^T K_x^{(s)} q_s}}. \tag{35}$$

Подставив это выражение в последние две формулы (3.72), будем иметь $D_s = g_s^T K_x^{(s)} g_s = 1$ и

$$x_s = K_x^{(s)} g_s = \frac{K_x^{(s)} q_s}{\sqrt{q_s^T K_x^{(s)} q_s}} \qquad (s = 1, \ldots, r). \tag{36}$$

Вычисление координат векторов $x_1, \ldots, x_r$ по этой формуле сводится к суммированию элементов соответствующих столбцов матриц $K_x^{(1)} = K_x$, $K_x^{(2)}, \ldots, K_x^{(r)}$ с изменением знаков некоторых из них и делению полученных сумм на корень квадратный из суммы всех элементов соответствующей матрицы $K_x^{(s)}$ с теми же изменениями знаков их элементов. Вследствие этого центроидный метод часто называют *методом простого суммирования*.

Непосредственное применение формулы (36), как уже было сказано, невозможно из-за неизвестности диагональных элементов матрицы K_x. Применяя первый способ преодоления этого затруднения, заменяют K_x в (36) найденной по результатам опытов оценкой $\hat{K}_y$ матрицы K_y. Определив первые r членов канонического разложения вектора Y, остальные члены включают в ошибку Z, пренебрегая корреляцией между координатами полученного таким путем вектора Z. При втором способе за оценки факторных дисперсий обычно принимают в первом приближении наибольшие модули элементов соответствующих строк и столбцов и применяют метод итераций. Выполнив вычисления по формуле (36) в первом приближении, определяют диагональные элементы матрицы K_x во втором приближении как суммы квадратов соответствующих координат векторов $x_1, \ldots, x_r$. Повторив вычисления по формуле (36), находят векторы $x_1, \ldots, x_r$ во втором приближении и т. д. Процесс последовательных приближений заканчивается, когда очередное приближение с достаточной точностью совпадет с предыдущим.

3.4. Вращение факторов. Учитывая, что факторы по определению должны иметь единичные дисперсии, приходим к выводу, что факторы определены с точностью

до произвольного ортогонального преобразования, т. е. произвольного вращения. Действительно, пусть $W = AV$. Тогда на основании (3.50) $K_w = AK_vA^T = AIA^T = AA^T$. Отсюда видно, что $K_w = I$ тогда и только тогда, когда $A^T = A^{-1}$, т. е. когда матрица A ортогональна. Произвольное ортогональное преобразование факторов приводит к новой системе факторов. Этим обычно пользуются для того, чтобы упростить интерпретацию результатов факторного анализа. Определив факторы, производят их вращение так, чтобы определенные координаты вектора Y имели наибольшие нагрузки на один фактор и близкие к нулевым на другие факторы.

Для первоначального определения факторов можно применять любой метод факторного анализа, в частности метод главных компонент *).

Ясно, что в преобразовании факторов путем вращения и в приведении их к «удобному» для анализа виду много субъективного. Поэтому разные исследователи могут интерпретировать одни и те же экспериментальные данные различно. В погоне за удобной для анализа формой факторов в приложениях, особенно в области психологии, часто преобразуют факторы так, что в итоге получаются коррелированные факторы.

При построении факторных моделей часто нормируют координаты наблюдаемого вектора Y, разделив их на оценки их средних квадратических отклонений. При этом ковариационная матрица K_y наблюдаемого вектора Y заменяется его корреляционной матрицей R_y.

Пример 7 **). Даны результаты 23 наблюдений 9-мерного случайного вектора (см. табл. на стр. 427).

Построить факторную модель, учитывая только те главные компоненты, которым соответствует 80% суммы всех собственных значений корреляционной матрицы.

В результате вычислений на ЭВМ по программе FACTO получены первые четыре собственных значения корреляционной матрицы

---

*) Метод факторного анализа, основанный на первоначальном определении факторов методом главных компонент, реализован в программе FACTO, которая использует для различных этапов факторного анализа стандартные подпрограммы CORRE (для вычисления выборочных средних, вторых моментов и коэффициентов корреляции), EIGEN (для определения собственных значений и собственных векторов), TRACE (для отбрасывания малых собственных значений), LOAD (для вычисления факторных нагрузок) и VARMX (для оптимального в известном смысле — *варимаксного* — вращения факторов) [106].

**) Этот пример заимствован из [94].

k	v_k^T								
1	7	7	9	7	15	36	60	15	24
2	13	18	25	15	13	35	61	18	30
3	9	18	24	23	12	43	62	14	31
4	7	13	25	36	11	12	63	26	32
5	6	8	20	7	15	46	18	28	15
6	10	12	30	11	10	42	27	12	17
7	7	6	11	7	15	35	60	20	25
8	16	19	25	16	13	30	64	20	30
9	9	22	26	24	13	40	66	15	32
10	8	15	26	30	13	10	66	25	34
11	8	10	20	8	17	40	20	30	18
12	9	12	28	11	8	45	30	15	19
13	11	17	21	30	10	45	60	17	30
14	9	16	26	27	14	31	59	19	17
15	10	15	24	18	12	29	48	18	26
16	11	11	30	19	19	26	57	20	30
17	16	9	16	20	18	31	60	21	17
18	9	8	19	14	16	33	67	9	19
19	7	18	22	9	15	37	62	11	20
20	8	11	23	18	9	36	61	22	24
21	6	6	27	23	7	40	55	24	31
22	10	9	26	26	10	37	57	27	29
23	8	10	26	15	11	42	59	20	28

наблюдаемого вектора: $\lambda_1 = 2{,}94989$, $\lambda_2 = 1{,}64368$, $\lambda_3 = 1{,}55515$ и $\lambda_4 = 1{,}06579$, сумма которых составляет 0,80161 от суммы всех девяти собственных значений (которая, очевидно, равна следу корреляционной матрицы, т. е. 9), и соответствующая четырехфакторная модель:

$$Y_1 = 9{,}30435 + 2{,}70412\,(0{,}05498V_1 + 0{,}07183V_2 - 0{,}05578V_3 + 0{,}85017V_4) + Z_1,$$

$$Y_2 = 12{,}60870 + 4{,}59978\,(0{,}29329V_1 - 0{,}39653V_2 - 0{,}35581V_3 + 0{,}60550V_4) + Z_2,$$

$$Y_3 = 23{,}00000 + 5{,}33427\,(0{,}05114V_1 - 0{,}82494V_2 + 0{,}15068V_3 + 0{,}32984V_4) + Z_3,$$

$$Y_4 = 18{,}00000 + 8{,}33393\,(0{,}74041V_1 - 0{,}41401V_2 + 0{,}24580V_3 + 0{,}13972V_4) + Z_4,$$

$$Y_5 = 12{,}86957 + 3{,}13781\,(-0{,}09091V_1 + 0{,}80662V_2 + 0{,}13525V_3 + 0{,}39228V_4) + Z_5,$$

$$Y_6 = 34{,}82608 + 9{,}29149\,(-0{,}68286V_1 - 0{,}21579V_2 - 0{,}44983V_3 - 0{,}20503V_4) + Z_6,$$

$$Y_7 = 54{,}00000 + 14{,}87826\,(0{,}86997V_1 + 0{,}18299V_2 - 0{,}34918V_3 + 0{,}08830V_4) + Z_7,$$

$$Y_8 = 19{,}39130 + 5{,}56563\,(0{,}03602V_1 - 0{,}05500V_2 + 0{,}91376V_3 - 0{,}15962V_4) + Z_8,$$

$$Y_9 = 25{,}13043 + 6{,}09249\,(0{,}80532V_1 - 0{,}32759V_2 + 0{,}00994V_3 - 0{,}02380V_4) + Z_9.$$

Первые слагаемые здесь представляют собой выборочные средние, а множители перед скобками — выборочные средние квадратические отклонения соответствующих координат вектора Y.

3.5. Применение метода п. 3.4.5. Чтобы избежать вращения факторов и сразу получить факторную модель с большим числом нулевых и близких к нулевым нагрузок на факторы, можно применить для нахождения канонического разложения вектора X метод п. 3.4.5. При этом, имея в виду, что порядок нумерации координат векторов может быть произвольным, можно по ходу вычислений устанавливать новую нумерацию этих координат с целью получить достаточно хорошую модель (по возможности с меньшей суммарной относительной величиной малых нагрузок).

Чтобы применить метод п. 3.4.5 для построения факторной модели, выберем в качестве векторов $q_1, \ldots, q_r$ единичные векторы первых r осей координат

$$q_s = [0 \ldots 0 \overset{s}{1} 0 \ldots 0]^T \qquad (s = 1, \ldots, r)$$

(буква s над единицей показывает, что она стоит на s-м месте). При таком выборе векторов $q_1, \ldots, q_r$ формула (36) дает

$$\begin{aligned} & x_{1p} = \frac{k_{1p}}{\sqrt{k_{11}}} \qquad (p = 1, \ldots, m), \\ & x_{s1} = \ldots = x_{s,\, s-1} = 0, \\ & x_{sp} = \frac{k_{sp} - x_{1s} x_{1p} - \ldots - x_{s-1,\, s} x_{s-1,\, p}}{\sqrt{k_{ss} - x_{1s}^2 - \ldots - x_{s-1,\, s}^2}} \\ & \qquad\qquad\qquad (p = s, \ldots, m;\ s = 2, \ldots, r), \end{aligned} \qquad (37)$$

где k_{pq} $(p, q = 1, \ldots, m)$ — элементы матрицы K_x.

Комбинируя формулы (37) с соответствующим изменением нумерации координат векторов, мы и получим факторную модель с наибольшим числом нулевых и малых факторных нагрузок, не прибегая к вращению факторов.

Чтобы найти критерий для установления новой нумерации координат векторов, заметим, что доля первого фактора V_1 в дисперсии p-й координаты вектора X равна отношению x_{1p}^2/k_{pp}. Поэтому естественно выбрать первую координату вектора X, а следовательно, и первый фактор V_1 так, чтобы суммарная доля первого фактора V_1 в дисперсиях координат $X_2, \ldots, X_m$ вектора X была ми-

нимальной. Тогда получаем условие

$$\varepsilon_1 = \sum_{p=2}^{m} \frac{x_{1p}^2}{k_{pp}} = \min. \qquad (38)$$

Принимая по очереди все координаты вектора X за первую и вычисляя каждый раз величину ε_1, выберем в качестве первой координаты ту, для которой ε_1 имеет наименьшее значение. Рассуждая так же, мы примем за s-ю ту из оставшихся $m-s+1$ координат вектора X, для которой величина

$$\varepsilon_s = \sum_{p=s+1}^{m} \frac{x_{sp}^2}{k_{pp} - x_{1p}^2 - \ldots - x_{s-1,p}^2} \quad (s = 2, \ldots, r) \qquad (39)$$

имеет наименьшее значение.

При выполнении вычислений по формулам (37) приходится, как и при применении других методов факторного анализа, или заменять K_x матрицей K_y, или применять метод итераций, задавая в первом приближении какие-нибудь грубо приближенные оценки диагональных элементов матрицы K_x.

Этот метод построения факторной модели (факторного анализа) представляет собой вариант метода *треугольной декомпозиции* [118]. Это название объясняется тем, что при этом методе в случае $r=m$ ковариационная матрица K_x представляется в виде произведения треугольной матрицы x, столбцами которой служат векторы $x_1, \ldots, x_m$, на ее транспонированную: $K_x = xx^T$.

3.6. Оценивание факторов. Во многих случаях, кроме построения факторной модели, необходимо также найти оценки факторов $V_1, \ldots, V_r$ в каждом из произведенных опытов. Для этого достаточно воспользоваться формулой (3.72). Подставив в нее выражение (35) вектора g_s, получим рекуррентные формулы

$$V_1 = \frac{q_1^T X^0}{\sqrt{q_1^T K_x q_1}},$$

$$V_s = \frac{q_s^T \left(X^0 - \sum_{p=1}^{s-1} V_p x_p \right)}{\sqrt{q_s^T K_x^{(s)} q_s}} \qquad (s = 2, \ldots, r).$$

Подставив сюда выражение $X^0 = Y + Z - m_x = Y + Z - m_y$ и заменив неизвестное m_y его оценкой $\overline{Y}$, получим

$$V_1 = \frac{q_1^T (Y - \overline{Y} + Z)}{\sqrt{q_1^T K_x q_1}},$$

$$V_s = \frac{q_s^T \left(Y - \overline{Y} - \sum_{p=1}^{s-1} V_p x_p + Z \right)}{\sqrt{q_s^T K_x^{(s)} q_s}} \quad (s = 2, \ldots, r).$$

В результате опытов получаются реализации $Y_1, \ldots, Y_n$ величины Y и вычисляется их выборочное среднее $\overline{Y}$ и оценки векторов $x_1, \ldots, x_r$. Соответствующие же реализации вектора ошибки Z остаются неизвестными. Чтобы получить оценки факторов в каждом из произведенных опытов, целесообразно воспользоваться регрессией факторов $V_1, \ldots, V_r$ на Y, т. е. принять за оценки факторов $V_1, \ldots, V_r$ их условные математические ожидания относительно наблюдаемого вектора Y. В результате, имея в виду, что $m_z = 0$, получим

$$\hat{V}_1 = \frac{q_1^T (Y - \overline{Y})}{\sqrt{q_1^T K_x q_1}},$$

$$\hat{V}_s = \frac{q_s^T \left(Y - \overline{Y} - \sum_{p=1}^{s-1} \hat{V}_p x_p \right)}{\sqrt{q_s^T K_x^{(s)} q_s}} \quad (s = 2, \ldots, r). \tag{40}$$

Другой метод получения линейной регрессии факторов на наблюдаемый вектор Y основан на применении уравнений (9.14) и (9.15), которые в данном случае имеют вид

$$g K_y = K_{vy}, \quad a = m_v - g m_y = -g m_y, \tag{41}$$

где K_{vy} — взаимная ковариационная матрица случайных векторов $V = [V_1 \ldots V_r]^T$ и Y. Оценка матрицы K_y находится непосредственно по экспериментальным данным. Остается найти оценку матрицы K_{vy}. На основании (32)

$$K_{vy} = M V Y^T = M V (m_x^T + V^T x^T + Z^T) = M V V^T x^T = K_v x^T = x^T,$$

где через x обозначена матрица, столбцами которой служат векторы $x_1, \ldots, x_r$. Таким образом, уравнения (41) принимают вид

$$g K_y = x^T, \quad a = -g m_y.$$

Решив эти уравнения и заменив m_y его оценкой $\overline{Y}$, получим $g = x^T K_y^{-1}$, $a = -x^T K_y^{-1} \overline{Y}$. После этого находим линейную регрессию вектора V на Y:

$$\hat{V} = gY + a = x^T K_y^{-1} (Y - \overline{Y}).$$

Отсюда получаем оценки факторов

$$\hat{V}_s = x_s^T K_y^{-1} (Y - \overline{Y}) \qquad (s = 1, \ldots, r). \tag{42}$$

Формулы (40) и (42) дают различные оценки факторов. Это различие объясняется тем, что при их выводе используются различные оценки регрессии факторов на наблюдаемый вектор.

§ 4. Модели распознавания

4.1. Математическая постановка задач распознавания. Как уже было сказано в п. 9.1.1, большое значение для практики имеет моделирование процессов распознавания. Задача распознавания состоит в том, что в результате наблюдения некоторого объекта из множества объектов, принадлежащих разным классам, определяется, какому из этих классов принадлежит наблюдаемый объект (например, по начертанию рукописной буквы определяется, какая буква написана; по наблюдаемому электромагнитному или звуковому сигналу определяется, каким источником порожден этот сигнал; по данным геологической разведки определяется, какие полезные ископаемые имеются в районе разведки, и т. п.). Ввиду большого разнообразия наблюдаемых и подлежащих распознаванию объектов их обычно объединяют абстрактным термином *образ* и говорят о *распознавании образов*. Такая терминология объясняется тем, что первыми задачами автоматического распознавания были задачи распознавания изображений.

Для построения модели процесса распознавания необходимо прежде всего наблюдаемые образы охарактеризовать некоторыми величинами, которые можно ввести в распознающую систему, например в ЭВМ. Эти величины могут иметь разные значения для различных образов, даже принадлежащих одному и тому же классу. Выбор числа и характера величин — *информативных признаков*, — определяющих образы, представляет собой первый этап построения модели распознавания. Этот вопрос должен

решаться конкретно для каждой задачи распознавания с учетом возможностей производить необходимые измерения. И лишь после этого можно дать математическую постановку данной задачи распознавания.

Дадим теперь общую математическую постановку задачи распознавания, которая может служить основой для построения моделей распознавания. Пусть x — совокупность всех величин — признаков, — характеризующих подлежащие распознаванию образы (в том числе и качественных признаков, принимающих только значения 0 и 1), y — номер класса образов. Задача состоит в том, чтобы по наблюденному значению x определить значение величины y.

В большей части задач практики наблюдаемые величины случайны и имеют различные распределения для разных классов образов, т. е. для разных значений y. Случайный характер наблюдаемых величин приводит к тому, что одно и то же значение x может наблюдаться при разных значениях y. Вследствие этого задача распознавания принципиально не может быть решена абсолютно точно и становится статистической. В таких случаях задача распознавания сводится к нахождению оценки $\hat{y}$ параметра y по наблюденному значению x случайной величины X.

Модель процесса распознавания представляет собой зависимость оценки $\hat{y}$ номера класса образов y от значения x наблюдаемой случайной величины X. Построение модели распознавания сводится к определению этой зависимости. Ясно, что эту зависимость следует определять так, чтобы число ошибок при применении модели распознавания было как можно меньше. Как и в случае регрессионных моделей, мы ограничимся здесь случаем, когда наблюдаемая величина X представляет собой конечномерный случайный вектор.

Предположим, что появления образов разных классов представляют собой события (несовместные и образующие полную группу). В этом случае номер y класса образов, которому принадлежит наблюдаемый образ, представляет собой реализацию дискретной случайной величины Y с возможными значениями $1, \ldots, N$, где N — число классов образов, подлежащих распознаванию. Пусть $p_1, \ldots, p_N$ — вероятности этих значений, т. е. априорные вероятности появления образов 1-го, ..., N-го классов, $f(x \mid y)$ —

условная плотность наблюдаемой величины X при данном значении y величины Y.

Очевидно, что задача распознавания может быть решена только в том случае, когда никакие две из плотностей $f(x|1), \ldots, f(x|N)$ не совпадают тождественно, т. е. когда величина X имеет различные распределения для образов разных классов. Это различие в условных распределениях величины X и служит признаком, по которому можно различать образы разных классов. Если какие-нибудь из плотностей $f(x|1), \ldots, f(x|N)$ одинаковы, то образы соответствующих классов неразличимы.

Вследствие случайности наблюдаемой величины X вырабатываемая моделью распознавания оценка $\hat{y}$ номера класса образов, как всегда в задачах математической статистики, представляет собой реализацию случайной величины $\hat{Y}$. Правильному решению соответствует совпадение $\hat{Y}$ с Y. Если $\hat{Y} \neq Y$, то модель распознавания принимает ошибочное решение. При этом возможны $N(N-1)$ различных видов ошибок в соответствии с тем, что модель может принять образ каждого из N классов за образ одного из $N-1$ других классов (каждому значению $y=1, \ldots, N$ величины Y может соответствовать любое из $N-1$ значений $\hat{Y}$, не совпадающих с y). Обозначим через α_{kl} условную вероятность события $\hat{Y}=k$ при $Y=l$,

$$\alpha_{kl} = P(\hat{Y}=k \mid Y=l) \qquad (k, l=1, \ldots, N).$$

Величины α_{kk} $(k=1, \ldots, N)$ представляют собой условные вероятности правильных решений при соответствующих значениях величины Y, а величины α_{kl} $(k, l=1, \ldots, N;\ k \neq l)$ — условные вероятности ошибок различных видов. Вероятность правильного решения (безусловная) определяется по формуле полной вероятности (1.14):

$$p = P(\hat{Y}=Y) = \sum_{k=1}^{N} p_k \alpha_{kk}. \qquad (43)$$

Вероятность ошибки (безразлично какого именно вида), очевидно, равна $q = 1-p$.

Построение детерминированной модели распознавания, очевидно, сводится к разбиению пространства значений величины X на N непересекающихся частей, которые мы обозначим через $A_1, \ldots, A_N$, и определению зависимости

$\hat{Y}$ от X формулой

$$\hat{Y}=k \text{ при } X\in A_k \qquad (k=1, \ldots, N).$$

Тогда модель распознавания будет принимать решение, что появился образ k-го класса при попадании величины X в область A_k $(k=1, \ldots, N)$.

Построение стохастической модели распознавания сводится к установлению зависимости от X вероятностей $\delta(1 \mid X), \ldots, \delta(N \mid X)$ отнесения наблюдаемого образа к первому, второму, и т. д., N-му классам. В этом случае модель распознавания будет по наблюдаемому значению x величины X вычислять вероятности $\delta(1 \mid x), \ldots$ $\ldots, \delta(N \mid x)$ и в соответствии с этими вероятностями путем статистического моделирования вырабатывать значение $\hat{y}$ оценки $\hat{Y}$.

4.2. Детерминированные модели распознавания. Решим сначала задачу построения детерминированной модели. Границы между областями $A_1, \ldots, A_N$ естественно определять так, чтобы вероятность правильного решения была максимальной. Такую модель будем называть *оптимальной* моделью распознавания. Вероятности α_{kl} $(k, l=1, \ldots, N)$ определяются в данном случае формулой

$$\alpha_{kl}=\int_{A_k} f(x \mid l)\, dx \qquad (k, l=1, \ldots, N). \tag{44}$$

Подставив в (43) выражение α_{kk} из (44), получим для вероятности правильного решения формулу

$$p=\sum_{k=1}^{n} p_k \int_{A_k} f(x \mid k)\, dx. \tag{45}$$

▶ Рассмотрим сначала случай двух классов образов, $N=2$. В этом случае область A_2 представляет собой дополнение области A_1, $A_2=\bar{A}_1$, и формула (45) принимает вид

$$p=p_1 \int_{A_1} f(x \mid 1)\, dx+p_2 \int_{\bar{A}_1} f(x \mid 2)\, dx.$$

Но

$$\int_{\bar{A}_1} f(x \mid 2)\, dx=1-\int_{A_1} f(x \mid 2)\, dx.$$

Следовательно,

$$p = p_2 + \int_{A_1} [p_1 f(x \mid 1) - p_2 f(x \mid 2)]\, dx. \quad (46)$$

Отсюда видно, что вероятность правильного решения будет максимальной, если за область A_1 принять всю область, в которой подынтегральная функция положительна. Действительно, если область A_1 представляет собой только часть области, в которой подынтегральная функция положительна, то интеграл в (46), а следовательно и вероятность правильного решения p, будет меньше. Если же включить в область A_1 часть области, в которой подынтегральная функция отрицательна, то интеграл в (46), а следовательно и вероятность p, тоже будет меньше. Таким образом, оптимальные области A_1 и A_2 определяются формулами*)

$$A_1 = \{x\colon\ p_1 f(x \mid 1) - p_2 f(x \mid 2) > 0\},$$
$$A_2 = \{x\colon\ p_1 f(x \mid 1) - p_2 f(x \mid 2) < 0\}. \quad (47)$$

Оптимальная граница между областями A_1 и A_2 определяется уравнением

$$p_1 f(x \mid 1) = p_2 f(x \mid 2).$$

На рис. 37 показана оптимальная граница — точка x_0 — между областями A_1 и A_2 в случае скалярной вели-

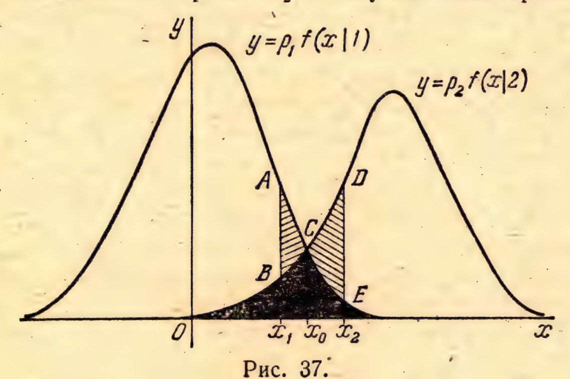


Рис. 37.

чины X. Вероятность ошибки $q = 1 - p$ равна площади зачерненного криволинейного треугольника. Если принять за границу между областями A_1 и A_2 точку x_1

*) Через $\{x\colon Q\}$, как всегда, обозначено множество значений x, удовлетворяющих условию Q.

левее точки x_0 или точку x_2 правее точки x_0, то вероятность ошибки увеличится на величину площади криволинейного треугольника ABC или соответственно CDE, отмеченных штриховкой. Следовательно, при смещении границы между областями A_1 и A_2 в любую сторону от точки x_0 вероятность ошибки q увеличивается и, следовательно, вероятность правильного решения p уменьшается.

Таким образом, оптимальная модель распознавания должна вырабатывать значение $\hat{y}$ оценки $\hat{Y}$ по следующему правилу:

$$\begin{aligned} \hat{y}=1 \quad \text{при} \quad \frac{f(x \mid 1)}{f(x \mid 2)} > \frac{p_2}{p_1}, \\ \hat{y}=2 \quad \text{при} \quad \frac{f(x \mid 1)}{f(x \mid 2)} < \frac{p_2}{p_1}. \end{aligned} \qquad (48)$$

Отсюда видно, что решающую роль в построении модели распознавания играет отношение условных плотностей величины X, соответствующих двум классам образов. Это отношение называется *отношением правдоподобия*. Очевидно, что вместо отношения правдоподобия $z = f(x \mid 1)/f(x \mid 2)$ можно взять любую строго возрастающую функцию этого отношения $\theta(z)$. ◀

Итак, *оптимальная модель распознавания должна вычислять некоторую строго возрастающую функцию отношения правдоподобия* θ/z *и сравнивать результат с величиной* $c=\theta(p_2/p_1)$, *называемой обычно порогом. При* $\theta(z) > c$ *принимается* $\hat{y}=1$. *При* $\theta(z) < c$ *принимается* $\hat{y}=2$. Этот алгоритм часто называется *решающим правилом*. Функция $\theta(z)$ называется *дискриминантной функцией*. Она выбирается так, чтобы вычисления были как можно проще.

Покажем, что найденная оптимальная модель распознавания является оптимальной и с точки зрения критерия максимума апостериорной вероятности, т. е. условной вероятности $P(Y=k \mid x)$ появления образа k-го класса при данном значении x величины X. Иными словами, модель принимает

$$\begin{aligned} \hat{y}=1 \quad \text{при} \quad P(Y=1 \mid x) > P(Y=2 \mid x), \\ \hat{y}=2 \quad \text{при} \quad P(Y=2 \mid x) > P(Y=1 \mid x). \end{aligned}$$

Для доказательства вычислим апостериорные вероятности $P(Y=k \mid x)$:

$$P(Y=k \mid x)=\frac{p_k f(x \mid k)}{p_1 f(x \mid 1)+p_2 f(x \mid 2)} \quad (k=1, 2).$$

Отсюда и из (47) видно, что $P(Y=1|x) > P(Y=2|x)$ при $x \in A_1$ и $P(Y=2|x) > P(Y=1|x)$ при $x \in A_2$.

▶ При произвольном числе классов образов N апостериорные вероятности $P(Y=k|x)$ определяются формулой

$$P(Y=k|x) = \frac{p_k f(x|k)}{\sum_{i=1}^{N} p_i f(x|i)} \qquad (k=1, \ldots, N)\ ^{*}).$$

Отсюда следует, что в оптимальной модели, работающей по критерию максимума апостериорной вероятности, области $A_1, \ldots, A_N$ определяются формулой

$$A_k = \{x\colon \max_h p_h f(x|h) = p_k f(x|k)\} \qquad (k=1, \ldots, N). \qquad (49)$$

Эта модель, очевидно, реализует и максимум вероятности правильного распознавания p, определяемой формулой (45), так как для нее формула (45) может быть переписана в виде

$$p = \int \max_h \{p_h f(x|h)\}\, dx. \qquad (50)$$

Пример 8. Рассмотрим случай распознавания образов двух классов при нормальных условных распределениях $f(x|1)$ и $f(x|2)$:

$$f(x|s) = \frac{1}{\sqrt{(2\pi)^r |K_s|}} \exp\left\{-\frac{1}{2}(x^T - m_s^T) K_s^{-1} (x - m_s)\right\}$$
$$(s=1, 2).$$

Ясно, что вычисления будут проще, если вместо отношения правдоподобия $z = f(x|1)/f(x|2)$ взять его удвоенный логарифм и отбросить в полученном выражении все слагаемые, не зависящие от x. Тогда получим функцию $\theta(z)$ в виде квадратного многочлена

$$\theta(z) = x^T(K_2^{-1} - K_1^{-1})\, x + 2(m_1^T K_1^{-1} - m_2^T K_2^{-1})\, x.$$

При этом порог c определится формулой

$$c = \theta\left(\frac{p_2}{p_1}\right) = 2 \ln \frac{p_2}{p_1} + \ln \frac{|K_1|}{|K_2|} + m_1^T K_1^{-1} m_1 - m_2^T K_2^{-1} m_2.$$

В частном случае совпадающих ковариационных матриц $K_1 = K_2 = K$ выражения для $\theta(z)$ и c принимают вид

$$\theta(z) = 2(m_1^T - m_2^T) K^{-1} x,$$
$$c = \theta\left(\frac{p_2}{p_1}\right) = 2 \ln \frac{p_2}{p_1} + m_1^T K^{-1} m_1 - m_2^T K^{-1} m_2.$$

---

*) Для вывода этой (и предыдущей) формулы достаточно подставить в формулу умножения плотностей (4.23) выражение плотности $f_2(y)$ дискретной случайной величины Y, затем вычислить $f_1(x)$ по формуле (4.8) и подставить найденные выражения $f(x, y)$ и $f_1(x)$ в (4.17).

В этом случае граница между областями A_1 и A_2 представляет собой гиперплоскость. В соответствии с этим решающее правило, по которому работает модель, называется *линейным*.

Найдем теперь условные вероятности ошибок: вероятность α_{12} принять образ второго класса за образ первого класса и вероятность α_{21} принять образ первого класса за образ второго класса. Для этого найдем условные распределения случайной величины

$$V = 2\ln\frac{f(X\mid 1)}{f(X\mid 2)} - \ln\frac{|K_1|}{|K_2|} - m_1^T K_1^{-1} m_1 + m_2^T K_2^{-1} m_2 = 2(m_1^T - m_2^T) K^{-1} X.$$

Так как V представляет собой линейную функцию вектора X с нормальными условными распределениями, то условные распределения V нормальны. Поэтому достаточно вычислить условные математические ожидания и дисперсии величины V. По формуле (3.11) находим

$$\mu_1 = M[V\mid 1] = 2(m_1^T - m_2^T) K^{-1} m_1,$$
$$\mu_2 = M[V\mid 2] = 2(m_1^T - m_2^T) K^{-1} m_2.$$

Условные дисперсии величины V найдем по формуле (3.50). Так как вектор X имеет одну и ту же ковариационную матрицу для образов обоих классов, то условные дисперсии величины V совпадают:

$$D = D[V\mid 1] = D[V\mid 2] = 4(m_1^T - m_2^T) K^{-1} K K^{-1} (m_1 - m_2) = 4(m_1^T - m_2^T) K^{-1}(m_1 - m_2).$$

По формуле (3.92) для вероятности попадания нормально распределенной случайной величины в интервал находим

$$\alpha_{12} = \frac{1}{\sqrt{2\pi D}}\int_c^{\infty} e^{\frac{-(v-\mu_2)^2}{2D}}\,dv = \frac{1}{2} - \Phi\left(\frac{c-\mu_2}{\sqrt{D}}\right),$$

$$\alpha_{21} = \frac{1}{\sqrt{2\pi D}}\int_{-\infty}^{c} e^{\frac{-(v-\mu_1)^2}{2D}}\,dv = \frac{1}{2} - \Phi\left(\frac{\mu_1 - c}{\sqrt{D}}\right).$$

Полная вероятность ошибки равна $q = p_1\alpha_{21} + p_2\alpha_{12}$, а вероятность правильного решения — $p = 1 - q = 1 - p_1\alpha_{21} - p_2\alpha_{12}$. Возможность точно определять вероятности ошибок двух видов при нормальном распределении вектора X является важным преимуществом линейных решающих правил.

4.3. Стохастические модели распознавания. Перейдем к стохастическим моделям распознавания.

▶ Условное распределение вероятностей $\delta(\hat{y}\mid x)$ ($\hat{y} = 1, \ldots, N$) при данном значении x величины X, в соответствии с которым стохастическая модель распознавания определяет оценку $\hat{Y}$ номера Y класса образов, называется *решающей функцией* или *рандомизированным правилом решения* модели распознавания *).

*) От английского слова «random» — «случайный».

Условные вероятности ошибок и правильных решений различных видов для стохастической модели распознавания определяются формулой

$$\alpha_{kl} = \int \delta(k \mid x) f(x \mid l)\, dx \qquad (k,\ l = 1, \ldots, N). \qquad (51)$$

Вероятность правильного распознавания на основании формулы полной вероятности (1.14) равна

$$p = \sum_{k=1}^{N} p_k \alpha_{kk} = \sum_{k=1}^{N} p_k \int \delta(k \mid x) f(x \mid k)\, dx = \\ = \int \sum_{k=1}^{N} p_k f(x \mid k)\, \delta(k \mid x)\, dx. \qquad (52)$$

Так как

$$0 \leqslant \delta(k \mid x) \leqslant 1, \qquad \sum_{k=1}^{N} \delta(k \mid x) = 1$$

при любом x, то подынтегральная функция в (52) представляет собой среднее взвешенное значение величин $p_k f(x \mid k)$ $(k = 1, \ldots, N)$, которое не может быть больше $\max\limits_k \{p_k f(x \mid k)\}$. Поэтому никакая стохастическая модель распознавания не может быть лучше оптимальной детерминированной модели, для которой согласно (50)

$$p = \int \max_k \{p_k f(x \mid k)\}\, dx.$$

Имея в виду определение (49) областей A_k, приходим к выводу, что вероятность правильного решения p достигает максимума тогда и только тогда, когда $\delta(k \mid x) = 1$ при $x \in A_k$ и $\delta(k \mid x) = 0$ при $x \notin A_k$ $(k = 1, \ldots, N)$, т. е. для оптимальной детерминированной модели. ◀

Таким образом, *оптимальная детерминированная модель является оптимальной среди всех возможных моделей, как детерминированных, так и стохастических.*

Только что доказанная теорема справедлива, конечно, лишь в том случае, когда при построении модели распознавания точно известны все вероятности $p_1, \ldots, p_N$ классов и условные плотности $f(x \mid 1), \ldots, f(x \mid N)$ величины X. Если модель распознавания строится при неполной информации о p_k и $f(x \mid k)$ $(k = 1, \ldots, N)$, когда они точно не известны, то стохастическая модель может оказаться лучше всех реализуемых детерминированных, так как оптимальная детерминированная модель остается в этом случае неизвестной и не может быть реализована.

4.4. Обучение моделей распознавания. Для построения модели распознавания в случае, когда вероятности p_k и плотности $f(x|k)$ $(k=1, \ldots, N)$ зависят от неизвестного параметра θ (в общем случае конечномерного векторного) или вообще неизвестны, их приходится оценивать по результатам опытов.

Если достаточно хорошие оценки неизвестного параметра θ можно получить непосредственно по наблюдаемому значению x вектора X (а это возможно только в случае, когда размерность вектора θ значительно меньше размерности вектора X), то можно включить в модель распознавания выработку соответствующих оценок вероятностей p_k и плотностей $f(x|k)$ и использование этих оценок в найденных решающих правилах.

Однако, как правило, хороших оценок параметра θ или непосредственно величин p_k и функций $f(x|k)$ по результатам наблюдения величины X получить невозможно. В таких случаях возникает задача *обучения* модели распознавания путем ввода в нее дополнительной информации, необходимой для выработки оценок неизвестных величин или непосредственно границ между областями $A_1, \ldots; A_N$. При этом в модель включается *алгоритм ее обучения*, т. е. формирования необходимых оценок на этапе обучения. Если во время обучения модели в нее вводятся только реализации случайной величины X (входного сигнала), то модель называется *самообучающейся*. Если же, кроме реализаций величины X, в модель вводится во время обучения информация о том, какому классу образов соответствует каждая реализация величины X, или какая-либо другая информация, которая может быть использована для оценивания неизвестных характеристик, то модель называется *обучаемой* (или *обучающейся*). При этом источник дополнительной информации называется *учителем* независимо от того, человек это или автоматическая распознающая система. Если учитель безошибочно определяет класс наблюдаемого образа, то он называется *идеальным*. Если же учитель может ошибаться, то он называется *реальным* [75—80].

Для оценивания неизвестных параметров или функций во время обучения модели можно использовать различные методы [25]. В частности, одним из наиболее эффективных методов является метод стохастических аппроксимаций (§§ 7.3 и 7.4).

4.5. Построение моделей распознавания без информации об априорных вероятностях. До сих пор мы предполагали априорные вероятности различных классов образов $p_1, \ldots, p_N$ известными. Однако во многих случаях они не известны и даже могут вообще не существовать, если образы различных классов при применении модели распознавания появляются не случайно. Во всех таких случаях оптимальной моделью распознавания образов двух классов считают такую модель, которая реализует минимум вероятности ошибки одного вида при данном значении вероятности ошибки другого вида. Условие минимума вероятности ошибки одного вида при заданной вероятности ошибки другого вида обычно называется *критерием Неймана—Пирсона.*

► Для нахождения решающего правила оптимальной модели распознавания по критерию Неймана—Пирсона можно применить метод неопределенных множителей Лагранжа. Согласно этому методу для нахождения условного минимума вероятности α_{21} при данном значении вероятности α_{12} надо сначала найти минимум величины $q' = \alpha_{21} + \lambda\alpha_{12}$, где λ — множитель Лагранжа, а потом определить λ так, чтобы вероятность α_{12} имела заданное значение. На основании (44) величина q' для детерминированной модели распознавания выражается формулой

$$q' = \int\limits_{A_2} f(x\,|\,1)\,dx + \lambda \int\limits_{A_1} f(x\,|\,2)\,dx.$$

Так как $A_2 = \bar{A}_1$ и вследствие этого

$$\int\limits_{A_2} f(x\,|\,1)\,dx = 1 - \int\limits_{A_1} f(x\,|\,1)\,dx,$$

то

$$q' = 1 - \int\limits_{A_1} [f(x\,|\,1) - \lambda f(x\,|\,2)]\,dx.$$

Отсюда совершенно так же, как в п. 4.2, находим оптимальные области A_1 и A_2, которые в данном случае зависят от неизвестного параметра λ:

$$A_1 = \{x\colon\ f(x\,|\,1) - \lambda f(x\,|\,2) > 0\},$$
$$A_2 = \{x\colon\ f(x\,|\,1) - \lambda f(x\,|\,2) < 0\}.$$

Сравнивая эти формулы с (47), приходим к выводу, что оптимальное правило решения отличается от найденного

в п. 4.2 только тем, что в выражении порога c вместо отношения вероятностей p_1/p_2 входит величина λ, $c=\theta(\lambda)$. Эта формула выражает c через величину λ, которая определяется из условия, чтобы вероятность α_{12} имела заданное значение. В задачах практики часто удается непосредственно определять порог c из условия равенства вероятности α_{12} заданному числу, что позволяет избежать промежуточного вычисления λ. ◀

Пример 9. В условиях примера 8 критерий Неймана—Пирсона дает для определения порога c уравнение

$$\frac{1}{\sqrt{(2\pi)^r |K_2|}} \int\limits_{A_1(c)} \exp\left\{-\frac{1}{2}(x^T - m_2^T) K_2^{-1}(x-m_2)\right\} dx = \alpha_{12},$$

где область интегрирования $A_1(c)$ определяется формулой

$$A_1(c) = \{x\colon\ x^T(K_2^{-1} - K_1^{-1})x + 2(m_1^T K_1^{-1} - m_2^T K_2^{-1})x > c\}.$$

В общем случае уравнение для c приближенно можно решить, вычисляя интеграл методом статистического моделирования для различных значений c. В частном случае одинаковых условных ковариационных матриц вектора X, $K_1 = K_2 = K$, т. е. в случае линейного решающего правила, уравнение для порога c на основании результатов примера 4 принимает вид

$$\Phi\left(\frac{c-\mu_2}{\sqrt{D}}\right) = \frac{1}{2} - \alpha_{12}.$$

Перейдем теперь к стохастическим моделям распознавания, обеспечивающим заданную вероятность α_{12}. Докажем, что никакая стохастическая модель при данной вероятности ошибки первого вида α_{12} не может иметь вероятность ошибки второго вида α_{21} меньшую, чем оптимальная детерминированная модель, основанная на критерии Неймана—Пирсона.

▶ Действительно, в этом случае

$$\int \delta(1\,|\,x) f(x\,|\,2)\,dx = \int\limits_{A_1} f(x\,|\,2)\,dx = \alpha_{12}, \tag{53}$$

где $A_1 = \{x\colon\ f(x\,|\,1) > \lambda f(x\,|\,2)\}$. Вероятность ошибки второго вида для детерминированной модели Неймана—Пирсона определяется формулой

$$\alpha_{21} = \int\limits_{A_2} f(x\,|\,1)\,dx,$$

а для стохастической модели — формулой

$$\alpha'_{21} = \int \delta(2 \mid x) f(x \mid 1)\, dx.$$

Следовательно,

$$\alpha'_{21} - \alpha_{21} = \int \delta(2 \mid x) f(x \mid 1)\, dx - \int_{A_2} f(x \mid 1)\, dx =$$

$$= \int_{A_1} \delta(2 \mid x) f(x \mid 1)\, dx - \int_{A_2} [1 - \delta(2 \mid x)] f(x \mid 1)\, dx =$$

$$= \int_{A_1} \delta(2 \mid x) f(x \mid 1)\, dx - \int_{A_2} \delta(1 \mid x) f(x \mid 1)\, dx.$$

Отсюда, имея в виду, что $f(x \mid 1) > \lambda f(x \mid 2)$ при $x \in A_1$ и $f(x \mid 1) < \lambda f(x \mid 2)$ при $x \in A_2$, получаем неравенство

$$\alpha'_{21} - \alpha_{21} \geqslant \lambda \left\{ \int_{A_1} \delta(2 \mid x) f(x \mid 2)\, dx - \right.$$

$$\left. - \int_{A_2} \delta(1 \mid x) f(x \mid 2)\, dx \right\} = \lambda \left\{ \int_{A_1} \delta(2 \mid x) f(x \mid 2)\, dx + \right.$$

$$\left. + \int_{A_1} \delta(1 \mid x) f(x \mid 2)\, dx - \int \delta(1 \mid x) f(x \mid 2)\, dx \right\} =$$

$$= \lambda \left\{ \int_{A_1} f(x \mid 2)\, dx - \int \delta(1 \mid x) f(x \mid 2)\, dx \right\}.$$

Но оба интеграла в фигурных скобках согласно (53) равны одной и той же вероятности α_{12}. Следовательно, $\alpha'_{21} \geqslant \alpha_{21}$, причем знак равенства достигается тогда и только тогда, когда $\delta(1 \mid x) = 1$ при $x \in A_1$ и $\delta(1 \mid x) = 0$ при $x \in A_2$, т. е. для оптимальной детерминированной модели. ◀

Таким образом, *оптимальная детерминированная модель Неймана — Пирсона является оптимальной среди всех моделей, как детерминированных, так и стохастических.*

Как и в п. 4.3, эта теорема доказана только для случая полностью известных распределений $f(x \mid 1)$ и $f(x \mid 2)$. Если они не известны или зависят от неизвестных параметров, то оптимальная модель Неймана — Пирсона не может быть реализована. В подобных случаях стохастическая модель может оказаться лучше всех детерминированных.

4.6. Проверка гипотез. Задачи распознавания родственны задачам проверки гипотез [49]. Основная задача теории проверки гипотез заключается в том, чтобы по результатам наблюдения случайной величины X, плотность вероятности которой $f(x|\theta)$ зависит от неизвестного параметра θ, принять или отвергнуть определенную гипотезу о значении θ. Если гипотеза состоит в том, что θ имеет определенное значение θ_0 и ни от каких других неизвестных параметров $f(x|\theta)$ не зависит, то гипотеза называется *простой*. Если же гипотеза состоит в том, что θ принадлежит некоторому множеству, то она называется *сложной*. В частности, гипотеза о том, что некоторые координаты векторного параметра θ имеют данные значения при неизвестных остальных координатах, будет сложной, так как она равноценна гипотезе, что θ принадлежит множеству всех векторов с данными значениями соответствующих координат.

Задача проверки простой гипотезы H_0, состоящей в том, что $\theta = \theta_0$, при условии, что θ может иметь другое определенное значение θ_1 (*конкурирующая* гипотеза H_1), очевидно, представляет собой задачу распознавания образов двух классов при

$$f(x|1) = f(x|\theta_1), \qquad f(x|2) = f(x|\theta_0)\ ^{*)}.$$

Отметим, что в этой задаче априорные вероятности p_1 и p_2 не существуют, так как θ — неслучайная величина, вследствие чего равенства $\theta = \theta_0$ и $\theta = \theta_1$ не являются событиями.

Решающее правило, для которого вероятность α_{12} отвергнуть гипотезу H_0, когда она верна, имеет данное значение ε, $\alpha_{12} = \varepsilon$, в теории проверки гипотез называется *правилом уровня* ε **).

Решающее правило данного уровня ε, для которого вероятность α_{11} отвергнуть гипотезу H_0, когда она неверна, максимальна, называется *наиболее мощным правилом уровня ε для гипотезы H_0 относительно гипотезы H_1*.

*) Мы относим значения наблюдаемой величины X, соответствующие интересующему нас значению θ_0 параметра θ (т. е. случаю, когда верна гипотеза H_0), ко второму классу образов, чтобы постановка задачи проверки гипотез совпала с постановкой решенной выше задачи распознавания.

**) В теории проверки гипотез детерминированное решающее правило, в соответствии с которым гипотеза H_0 принимается или отвергается, называется *критерием*. Мы предпочитаем более общий и более современный термин *решающее правило*.

Если одно и то же решающее правило оказывается наиболее мощным при всех возможных значениях θ_1 (при всех конкурирующих гипотезах H_1), то оно называется *равномерно наиболее мощным правилом уровня* ε *для гипотезы* H_0 *относительно множеств допустимых гипотез.*

Современная теория проверки гипотез заложена трудами Неймана и Е. Пирсона [67, 68], которые первыми решили задачу нахождения наиболее мощного решающего правила данного уровня. Именно поэтому критерий максимума вероятности α_{11} (минимума α_{21}) при данной вероятности α_{12} называется критерием Неймана—Пирсона.

Ясно, что решающее правило оптимальной модели распознавания образов двух классов, основанное на критерии Неймана—Пирсона, является наиболее мощным правилом уровня $\varepsilon=\alpha_{12}$ для гипотезы H_0 относительно H_1. Случаи, когда существует равномерно наиболее мощное решающее правило, встречаются очень редко.

Пример 10. Найденная в примере 9 оптимальная модель распознавания решает задачу проверки гипотезы о том, что неизвестное математическое ожидание m и ковариационная матрица K случайного вектора X равны m_2 и K_2 (при конкурирующей гипотезе $m=m_1$, $K=K_1$). Решающее правило этой модели является наиболее мощным правилом уровня $\varepsilon=\alpha_{12}$. Согласно этому правилу гипотеза $m=m_2$, $K=K_2$ принимается, если $\theta(z) < c$, и отвергается, если $\theta(z) > c$. При известных совпадающих ковариационных матрицах K_1 и K_2, $K_1=K_2=K$, найденное в примере 5 оптимальное линейное решающее правило является наиболее мощным правилом уровня $\varepsilon=\alpha_{12}$ для гипотезы $m=m_2$ относительно гипотезы $m=m_1$.

Если модель распознавания строится в случае неизвестных плотностей $f(x \mid k)$ или известных с точностью до неопределенного параметра θ, то, так же как в случае моделей п. 4.2, возникает задача обучения модели, о которой сказано в конце п. 4.4.

4.7. Последовательные модели распознавания. В изложенной теории построения моделей распознавания вектор X может представлять собой или однократно наблюдаемую случайную величину, или совокупность результатов n последовательных наблюдений некоторого случайного вектора U меньшей размерности. В последнем случае $X=[U_1^T \ \dots \ U_n^T]^T$, где $U_1, \dots, U_n$ — результаты наблюдений вектора U, и рассмотренные модели распознавания вырабатывают решение только после того, как произведены все n наблюдений. Однако можно построить и такую модель распознавания двух классов образов, которая после каждого наблюдения или принимает определенное решение о классе наблюдаемого образа, или принимает решение сделать еще одно наблюдение. Такой

метод решения задачи распознавания, называемый *последовательным анализом*, разработан Вальдом [14] *).

Пусть $f_k(x|1)$ и $f_k(x|2)$ — условные плотности вектора X для образов двух классов для первых k наблюдений ($k=1, 2, \ldots$). Размерность вектора X в этом случае, очевидно, равна kp, где p — размерность вектора U. Пусть $\theta_k(z)$, $z = f_k(x|1)/f_k(x|2)$, — дискриминантная функция для первых k наблюдений. Метод последовательного анализа состоит в том, что, задав обе условные вероятности ошибки α_{12} и α_{21}, на каждом шаге определяют два порога c_{1k} и c_{2k}, $c_{1k} < c_{2k}$ ($k=1, 2, \ldots$). Если после k-го наблюдения $\theta_k(z) > c_{2k}$, то принимается $\hat{y}=1$. Если $\theta_k(z) < c_{1k}$, то принимается $\hat{y}=2$. Если же $c_{1k} < \theta_k(z) < c_{2k}$, то производится $(k+1)$-е наблюдение вектора U **). Ясно, что в этом случае модель примет образ второго класса за образ первого класса, если для образа второго класса получится $\theta_k(z) > c_{2k}$, а образ первого класса за образ второго класса, если для образа первого класса получится $\theta_k(z) < c_{1k}$. Следовательно, условные вероятности ошибок двух видов выражаются через пороги c_{1k} и c_{2k} формулами

$$\alpha_{12} = \int\limits_{\theta_k(z) > c_{2k}} f(x|2)\,dx, \quad \alpha_{21} = \int\limits_{\theta_k(z) < c_{1k}} f(x|1)\,dx. \qquad (54)$$

При заданных вероятностях α_{12} и α_{21} эти соотношения служат уравнениями для определения порогов c_{1k} и c_{2k}. Эти уравнения в общем случае можно приближенно решить, вычисляя интегралы для ряда значений c_{1k} и c_{2k} путем статистического моделирования (§ 8.4).

Число наблюдений, после которых последовательная модель распознавания принимает определенное решение о классе образа, представляет собой случайную величину. Покажем, что при любых вероятностях α_{12} и α_{21}, с которыми задача распознавания решается оптимальной

*) Изложение современной теории последовательного анализа, из которой первоначальные методы Вальда вытекают как частный случай, дано в [124].

**) При этом, конечно, образ, предъявленный для распознавания, не должен изменяться в течение всего времени наблюдений. Иными словами, многократно наблюдаемая случайная величина U должна характеризовать один и тот же образ (иметь одно и то же распределение) во время всех наблюдений. В случае проверки простых гипотез все наблюдения вектора U должны производиться при одном и том же значении неизвестного параметра θ.

моделью Неймана—Пирсона при каком-нибудь конечном числе наблюдений n, математическое ожидание числа наблюдений, необходимых для последовательного распознавания с теми же вероятностями α_{12} и α_{21}, всегда меньше n.

▶ Предположим, что при n наблюдениях вектора U оптимальная модель Неймана—Пирсона решает задачу распознавания с вероятностями ошибок α_{12} и α_{21}. Тогда

$$\int\limits_{\theta_n(z)>c} f_n(x\,|\,2)\,dx = \alpha_{12}, \qquad \int\limits_{\theta_n(z)<c} f_n(x\,|\,1)\,dx = \alpha_{21}.$$

Пороги c_{1n} и c_{2n} на n-м шаге последовательного распознавания с теми же вероятностями ошибок α_{12} и α_{21} определяются уравнениями (54) при $k=n$. Сравнивая последние равенства с (54) при $k=n$, получаем $c_{1n}=c_{2n}=c$. Вследствие этого последовательная модель распознавания с вероятностью 1 примет определенное решение после не больше чем n наблюдений (событие $\theta_n(z)=c$ практически невозможно; поэтому $(n+1)$-го наблюдения заведомо не будет). Однако последовательная модель может принять определенное решение и раньше, после $k<n$ наблюдений, и вероятность этого события не равна нулю. Следовательно, математическое ожидание случайного числа наблюдений, при котором последовательная модель распознавания примет определенное решение, меньше n. ◀

Пример 11. Предположим, что в условиях примера 8 вектор X представляет собой совокупность результатов k независимых наблюдений случайного вектора U, $X=[U_1^T \dots U_k^T]^T$. В этом случае выражение примера 8 для дискриминантной функции принимает вид

$$\theta_k(z) = \sum_{i=1}^{k} [u_i^T (L_2^{-1} - L_1^{-1})\, u_i + (h_1^T L_1^{-1} - h_2^T L_2^{-1})\, u_i],$$

где h_1, h_2 — условные математические ожидания, а L_1, L_2 — условные ковариационные матрицы вектора U для образов двух классов.

В частном случае, когда $L_1=L_2=L$, формула для $\theta_k(z)$ принимает вид

$$\theta_k(z) = \sum_{i=1}^{k} (h_1^T - h_2^T)\, L^{-1} u_i$$

и формулы для условных математических ожиданий и дисперсий величины

$$V = \theta_k(Z) = \sum_{i=1}^{k} (h_1^T - h_2^T)\, L^{-1} U_i$$

дают

$$\mu_{1k}=k\,(h_1^T-h_2^T)\,L^{-1}h_1,\quad \mu_{2k}=k\,(h_1^T-h_2^T)\,L^{-1}h_2,$$
$$D_k=k\,(h_1^T-h_2^T)\,L^{-1}\,(h_1-h_2).$$

Полагая для краткости $(h_1^T-h_2^T)\,L^{-1}h_1=\lambda_1$, $(h_1^T-h_2^T)\,L^{-1}h_2=\lambda_2$, будем иметь

$$\mu_{1k}=k\lambda_1,\quad \mu_{2k}=k\lambda_2,\quad D_k=k\,(\lambda_1-\lambda_2).$$

Уравнения (54) для определения порогов c_{1k} и c_{2k} в этом случае имеют вид

$$\Phi\left(\frac{k\lambda_1-c_{1k}}{\sqrt{k\,(\lambda_1-\lambda_2)}}\right)=\frac{1}{2}-\alpha_{21},\quad \Phi\left(\frac{c_{2k}-k\lambda_2}{\sqrt{k\,(\lambda_1-\lambda_2)}}\right)=\frac{1}{2}-\alpha_{12}.$$

Обозначив аргументы функции Лапласа, при которых справедливы эти равенства, соответственно ν_1 и ν_2, получим

$$c_{1k}=k\lambda_1-\nu_1\sqrt{k\,(\lambda_1-\lambda_2)},\quad c_{2k}=k\lambda_2+\nu_2\sqrt{k\,(\lambda_1-\lambda_2)}.$$

Следовательно,

$$c_{2k}-c_{1k}=(\nu_1+\nu_2)\sqrt{k\,(\lambda_1-\lambda_2)}-k\,(\lambda_1-\lambda_2).$$

Отсюда видно, что число наблюдений n, при котором $c_{1n}\leqslant c_{2n}$, определяется неравенством

$$n\geqslant\frac{(\nu_1+\nu_2)^2}{\lambda_1-\lambda_2}.$$

Наименьшее целое число n (всегда положительное вследствие того, что $\lambda_1>\lambda_2$), удовлетворяющее этому условию, и будет максимальным возможным значением случайного числа наблюдений, необходимых для того, чтобы последовательная модель распознавания приняла определенное решение. Математическое ожидание числа наблюдений для последовательной модели распознавания будет всегда меньше n.

Последовательный анализ можно считать началом теории управления экспериментальными исследованиями. В последовательном анализе Вальда управление состоит в том, что в зависимости от результатов проведенных опытов на каждом шаге решается, продолжать опыты или прекратить их и принять определенное решение о принадлежности результатов наблюдений определенному классу. В более общих и сложных задачах управление должно включать и решение о том, какие именно экспериментальные исследования следует провести, в зависимости от результатов обработки уже проведенных опытов. В таком плане теория управления экспериментальными исследованиями сейчас еще только начинает развиваться. Обобщение методов последовательного анализа содержится в статистической теории решений, основы которой также были заложены Вальдом [15].

§ 5. Модели принятия решений

5.1. Задача принятия решений. Все задачи, решаемые моделями любого назначения, можно рассматривать как задачи принятия решений. Так, например, регрессионная модель по данному значению величины x вырабатывает оценку величины Y, т. е. принимает решение о том, какое значение следует приписать величине Y. Модель распознавания по наблюдаемому значению x величины X определяет, к какому классу следует отнести соответствующий этому x образ, т. е. принимает решение о том, какому классу принадлежит наблюдаемый образ. Точно так же любую задачу оценивания можно рассматривать как задачу решения о том, какие значения следует приписать неизвестным статистическим характеристикам, имея в своем распоряжении результаты наблюдений. Эти соображения привели Вальда к разработке основ общей статистической теории решений [15]. В дальнейшем эта теория начала распространяться в направлении разработки методов обоснования решений, принимаемых людьми во всех областях деятельности человека, т. е. методов построения моделей процессов обоснования и принятия любых решений. Сейчас теория решений интенсивно развивается и составляет один из важнейших разделов общей теории управления [4, 21, 52, 61, 80].

В простейших задачах решения принимаются при полной информации о возможных последствиях различных решений. Так, например, при построении регрессионной модели в случае известного совместного распределения величин X и Y полностью известно распределение ошибок модели (отклонений от регрессии). При построении модели распознавания в случае известных условных плотностей $f(x|k)$ $(k=1, \ldots, N)$ можно вычислить вероятности ошибок всех видов, а если известны также и априорные вероятности появления образов различных классов $p_1, \ldots, p_N$, то и полную вероятность ошибки q.

Однако в большей части задач практики решения приходится принимать при неполной информации. Так, например, при построении регрессионной модели в случае, когда распределение величин X и Y не известно или зависит от неизвестных параметров, распределение ошибки модели остается в той или иной степени неопределенным. Точно так же при построении модели распознавания в случае, когда p_k и $f(x|k)$ $(k=1, \ldots, N)$ не

известны или зависят от неизвестных параметров (при проверке сложных гипотез), точно вычислить вероятности ошибок разных видов невозможно.

Еще больше неопределенности в задачах принятия решений людьми, участвующими в управлении. В таких задачах возможные последствия различных решений большей частью трудно, а иногда невозможно оценить. Решение приходится принимать в условиях неопределенности. Задача науки при этом состоит в том, чтобы определить наиболее разумные варианты решений, связанные с наименьшим риском неудач и потерь. Мы говорим здесь о наиболее разумных, а не об оптимальных решениях потому, что решения, как правило, приходится оценивать с различных точек зрения. Решения, оптимальные с одной точки зрения, могут оказаться плохими с другой точки зрения. Поэтому выбор определенного решения практически всегда связан с необходимостью компромисса между разными, подчас противоречивыми требованиями.

5.2. Риск и функция потерь. Для построения теории решений необходимо прежде всего ввести какую-то меру качества решения. Пусть x — совокупность всех величин, характеризующих исходные данные для принятия решения (т. е. всю информацию, которой мы располагаем для принятия решения — входной сигнал модели решения), z — совокупность всех величин, характеризующих решение (выходной сигнал модели решения). Качество решения можно характеризовать функцией двух переменных $r(z \mid x)$, определяющей потери (или проигрыш), к которым приводит решение z при данном x. Эту функцию обычно называют *риском, связанным с решением z при данном x* (*стоимостью* или *ценой* решения z при данном x) *).

Пример 12. В случае оценивания регрессии путем минимизации среднего квадрата ошибки при данном x качество модели характеризуется квадратом модуля ошибки $|z-Y|^2$. Определив риск формулой

$$r(z \mid x) = M[|z-Y|^2 \mid x] = \int |z-y|^2 f(y \mid x)\, dx,$$

можно свести задачу п. 9.2.1 к минимизации риска.

---

*) Вместо риска часто вводят *полезность* решения z при данном x — функцию $u(z \mid x)$, характеризующую выигрыш (прибыль), к которому приводит решение z при данном x. Однако с математической точки зрения это одно и то же, так как при данной полезности решения $u(z \mid x)$ всегда можно определить риск, положив $r(z \mid x) = r_0 - u(z \mid x)$, где r_0 — произвольное число, и таким путем свести задачу максимизации полезности к задаче минимизации риска.

П р и м е р 13. В случае распознавания по максимуму апостериорной вероятности с правильным решением можно связать нулевые потери, а ошибке любого вида приписать потери, равные 1. Определив таким путем функцию потерь

$$l(y, z) = \begin{cases} 0 & \text{при} \quad z = y, \\ 1 & \text{при} \quad z \neq y, \end{cases}$$

можно определить риск формулой

$$r(z \mid x) = M[l(Y, z) \mid x] = 1 - \frac{p_k f(x \mid k)}{\sum_{i=1}^{N} p_i f(x \mid i)}.$$

Тогда задача построения оптимальной детерминированной модели распознавания сведется к минимизации риска.

П р и м е р 14. Задача оценивания неизвестного параметра θ методом максимума правдоподобия § 7.2 сводится к минимизации риска, определяемого формулой

$$r(z \mid u) = r_0 - g(u \mid z),$$

где $g(u \mid \theta)$ — функция правдоподобия, а r_0 — произвольная постоянная.

Способы определения риска в примерах 12 и 13 представляют собой частные случаи общего способа определения риска в задачах оценивания случайных величин. Этот общий способ состоит в том, что для любых возможных значений y, z случайной величины Y и ее оценки Z потери оцениваются значением некоторой функции $l(y, z)$. Эта функция называется *функцией потерь*. При заданной функции потерь риск определяется как условное математическое ожидание функции потерь при данных значениях x и z:

$$r(z \mid x) = M[l(Y, z) \mid x] = \int l(y, z) f(y \mid x)\, dx. \qquad (55)$$

В примере 12 $l(y, z) = |z - y|^2$, а в примере 13 $l(y, z) = 1 - \delta_{yz}$ $(y, z = 1, \ldots, N)$. Другими примерами применяемых в задачах практики функций потерь могут служить $l(y, z) = 1 - \exp\{-k^2 |z - y|^2\}$, $l(y, z) = 1 - 1/(1 + k^2 |z - y|^2)$, $l(y, z) = l_{ij}$ при $y = i$, $z = j$ $(i, j = 1, \ldots, N)$. Если за функцию потерь принять индикатор некоторого множества, $l(y, z) = 1_{A(z)}(y)$, то риск будет равен условной вероятности попадания случайной величины Y на это множество при данных x и z:

$$r(z \mid x) = \int 1_{A(z)}(y) f(y \mid x)\, dy = \int_{A(z)} f(y \mid x)\, dy = P(Y \in A(z) \mid x).$$

При этом чаще всего за множество $\overline{A(z)}$ принимают круг некоторого радиуса a с центром в точке z в пространстве значений величины Y (интервал $(z-a,\ z+a)$ в случае скалярной величины Y). При этом риск определяется формулой

$$r(z \mid x)=P(|Y-z|>a \mid x).$$

Формулу (55) можно формально распространить и на задачи оценивания неслучайного параметра $y=\theta$. Для этого достаточно подставить в (55) выражение условной плотности $f(y \mid x)$ из формулы умножения плотностей (4.23) и воспользоваться формулой (4.8). Тогда получим

$$r(z \mid x)=\frac{\int l(y,\ z)\, f_2(y)\, f_1(x \mid y)\, dy}{\int f_2(y)\, f_1(x \mid y)\, dy}.$$

В случае равномерного распределения Y в некоторой области B функция $f_2(y)$ постоянна в области B, вследствие чего предыдущая формула принимает вид

$$r(z \mid x)=\frac{\int\limits_B l(y,z)\, f_1(x \mid y)\, dy}{\int\limits_B f_1(x \mid y)\, dy}.$$

Если функция $f(x \mid y)$ интегрируема по y по всему пространству значений величины Y, то можно формально перейти к пределу, когда область B неограниченно расширяется и стремится совпасть со всем пространством значений величины Y. В результате получаем

$$r(z \mid x)=\frac{\int l(y,z)\, f_1(x \mid y)\, dy}{\int f_1(x \mid y)\, dy}. \tag{56}$$

Эта предельная формула имеет смысл только тогда, когда y представляет собой неслучайный параметр, так как никакая случайная величина не может быть равномерно распределенной на всем пространстве ее значений. Поэтому формулой (56) обычно определяют риск в задачах оценивания неслучайных параметров (или параметров, относительно которых не известно, случайны они или нет). Положив, в частности, $l(y,\ z)=r_0-\delta(z-y)$, получаем $r(z \mid x)=r_0-c(x)\, f_1(x \mid z)$, где через $c(x)$ обозначена

обратная величина интеграла в знаменателе формулы (56). Таким образом, взяв в качестве функции потерь отрицательную δ-функцию $-\delta(z-y)$, можно свести оценивание параметров по методу максимума правдоподобия к минимизации риска. Постоянное слагаемое r_0 и множитель $c(x)$, зависящий только от значения x наблюдаемой величины X, при этом не играют никакой роли, и r_0 вводится только для того, чтобы формально не получать отрицательных значений риска (с математической точки зрения без потери общности всегда можно принять $r_0=0$, так же как и в примере 14).

В задачах решения, в которых целью не является выработка оценки какой-нибудь случайной или неслучайной величины, риск нельзя определить с помощью функции потерь так, как это делается в задачах оценивания. Поэтому в таких задачах приходится непосредственно определять риск $r(z|x)$ как функцию исходных данных x и решения z. Сейчас разрабатываются методы определения риска (или полезности) для различных задач принятия решений с учетом индивидуальных предпочтений и черт характера людей, которые должны принимать решения [61, 80].

Если риск представляет собой скалярную величину (задача решения *однокритериальна*), то имеет смысл говорить об *оптимальном решении*, минимизирующем риск. Однако во многих задачах практики при выработке решения приходится руководствоваться многими критериями, которые невозможно свести к одному скалярному критерию (*многокритериальные* задачи решения). В таких случаях риск представляет собой векторную функцию и об оптимальных решениях, строго говоря, речи быть не может. Действительно, в общем случае не существует решения, для которого все координаты векторного риска одновременно имеют минимальные возможные значения. Тем не менее на практике часто говорят об оптимизации по векторным критериям, подразумевая под этим выработку приемлемого решения, достаточно хорошего с точки зрения всех скалярных критериев. Существует ряд методов для выработки таких приемлемых решений [29]. Все они в той или иной степени включают определение решений, оптимальных с точки зрения каждого скалярного критерия по отдельности. Поэтому мы рассмотрим здесь кратко основные принципы построения моделей решения, основанных на минимизации какого-нибудь

одного критерия (скалярного риска). Методы увязки таких частных оптимальных решений и нахождения приемлемых решений с учетом нескольких критериев увели бы нас далеко за рамки этой книги, и мы их рассматривать не будем.

5.3. Оптимальные решения. Рассмотрим сначала задачу оптимизации решения, не ограничивая класс возможных решений. При этом мы должны будем рассматривать не только детерминированные, но и стохастические модели решения. Стохастическая модель вырабатывает решение, представляющее собой реализацию случайной величины Z, в соответствии с некоторым распределением. Это распределение, конечно зависящее от входного сигнала x как от параметра, называется *решающей функцией* стохастической модели решения. Поскольку мы рассматриваем здесь только конечномерные случайные величины, можно считать, что решающая функция задается плотностью $\delta(z \mid x)$. Детерминированная модель, вырабатывающая определенное решение $z=\varphi(x)$, представляет собой частный случай стохастической модели, когда $\delta(z \mid x) = \delta(z-\varphi(x))$.

Для стохастической модели риск как функция случайного решения Z сам представляет собой случайную величину $r(Z \mid x)$. Поэтому оптимальной стохастической моделью называется модель, минимизирующая математическое ожидание риска (т. е. средний риск) при данном x

$$\rho(\delta \mid x) = M[r(Z \mid x) \mid x] = \int r(z \mid x)\, \delta(z \mid x)\, dz. \qquad (57)$$

Без всяких выкладок ясно, что *если существует такая функция $z=\varphi(x)$, при которой риск $r(\varphi(x) \mid x)$ имеет наименьшее возможное значение $r(\varphi(x) \mid x) = \inf\limits_z r(z \mid x)$ при любом x, то детерминированное решение $z=\varphi(x)$ будет оптимальным среди всех возможных решений и никакая стохастическая модель не может быть лучше детерминированной, принимающей $z=\varphi(x)$ при любом x.* К этому же выводу приходим и математически. Если $r(\varphi(x) \mid x) = \inf\limits_z r(z \mid x)$, то $r(z \mid x) \geqslant r(\varphi(x) \mid x)$ при любых x и z, вследствие чего

$$\rho(\delta \mid x) \geqslant \int r(\varphi(x) \mid x)\, \delta(z \mid x)\, dz = r(\varphi(x) \mid x)$$

для любой решающей функции $\delta(z \mid x)$.

Так как условное математическое ожидание $m_y(x) = M[Y|x]$ представляет собой функцию, для которой средний квадрат ошибки $M[|z-Y|^2|x]$ достигает точной нижней грани при всех x (п. 3.2.4),

$$M[|m_y(x)-Y|^2|x] = \inf_z M[|z-Y|^2|x],$$

то из доказанной теоремы следует, что *никакая стохастическая модель прогноза или оценивания случайной величины Y по результатам наблюдения x не может быть лучше оптимальной детерминированной модели, работающей по регрессии.* Этот вывод дополняет результаты п. 9.2.1.

Точно так же в задаче распознавания п. 4.5 существует функция

$$\hat{y}(x) = k \quad \text{при} \quad p_k f(x|k) = \max_l \{p_l f(x|l)\} \quad (k = 1, \ldots, N),$$

минимизирующая риск примера 13 при всех x. Поэтому не существует стохастической модели распознавания, лучшей, чем оптимальная детерминированная модель.

Если при любом x существует непустое множество

$$D_x = \{z\colon\ r(z|x) = \inf_{z'} r(z'|x)\},$$

то любое распределение вероятностей, полностью сосредоточенное на множестве D_x, будет решающей функцией стохастической оптимальной системы, так как для любой плотности вероятности $\delta(z|x)$, равной 0 вне D_x,

$$\rho(\delta|x) = \int_{D_x} r(z|x)\,\delta(z|x)\,dx = \inf r(z|x).$$

В этом случае существует бесчисленное множество оптимальных систем, как детерминированных, так и стохастических, и все они равноценны, средний риск $\rho(\delta|x)$ для них имеет одно и то же значение.

Если точная нижняя грань $\inf r(z|x)$ не достигается ни при каком конечном z в области возможных значений x, то оптимального решения не существует. Однако по определению точной нижней грани всегда существуют значения z, при которых риск $r(z|x)$ как угодно близок к точной нижней грани. Поэтому для произвольного $\varepsilon > 0$ существует множество значений z, для которых риск $r(z|x)$ отличается от своей нижней грани при данном x

не больше чем на ε:

$$D_x^\varepsilon = \{z\colon\ r(z\,|\,x) - \inf_{z'} r(z'\,|\,x) < \varepsilon\}.$$

Для любой плотности $\delta(z\,|\,x)$, равной 0 вне области D_x^ε,

$$\rho(\delta\,|\,x) = \int\limits_{D_x^\varepsilon} r(z\,|\,x)\,\delta(z\,|\,x)\,dz < \inf r(z\,|\,x) + \varepsilon.$$

Таким образом, всегда существует бесконечное множество решений, как детерминированных, так и стохастических, как угодно близких к оптимальному (в том смысле, что средний риск для них как угодно близок к своей точной нижней грани). Все эти решения с практической точки зрения не отличаются от оптимального. Поэтому с практической точки зрения всегда существует множество оптимальных решений, даже тогда, когда математическая задача оптимизации не имеет решения.

5.4. Оптимальные решения в данном классе функций. Если оптимальное детерминированное решение ищется в ограниченном классе функций Ψ, то в общем случае оно не может быть оптимальным при каждом данном значении x. В таких случаях, как и в п. 9.2.1, приходится ставить задачу минимизации среднего риска в данной области изменения x. Так же как и в п. 9.2.1, приходим к выводу, что при любом методе осреднения $r(z\,|\,x)$ можно считать x значением случайной величины X и минимизировать $Mr(z(X)\,|\,X)$.

Пример 15. В условиях примера 12 оптимальное детерминированное решение $z(x)$ в данном классе функций Ψ определяется из условия минимума величины

$$Mr(z(X)\,|\,X) = \int f_1(x)\,dx \int |z(x) - y|^2\, f(y\,|\,x)\,dx.$$

5.5. Принятие решений в условиях неопределенности. Оптимальные или близкие к оптимальным модели решения, минимизирующие средний риск, могут быть реализованы или хотя бы исследованы только в том случае, когда риск $r(z\,|\,x)$ представляет собой известную функцию. Однако во многих задачах риск невозможно определить как известную функцию, так как при одних и тех же значениях x и z последствия решения z точно определить (прогнозировать) невозможно, вследствие того что они зависят от многих неизвестных факторов, как случайных, так и неслучайных. Если действие всех этих факторов можно охарактеризовать в каждом конкретном случае

конечным набором чисел, то риск иногда удается представить как известную функцию x, z и неизвестного векторного параметра θ (параметризовать риск). Обозначим его через $r(z|x, \theta)$. Тогда и средний риск при данной решающей функции $\delta(z|x)$ будет зависеть от θ:

$$\rho(\delta|x, \theta) = \int r(z|x, \theta)\,\delta(z|x)\,dz. \tag{58}$$

Минимизация среднего риска в этом случае даст оптимальную решающую функцию $\delta_\theta(z|x)$, тоже зависящую от θ. Вследствие этого такая оптимальная решающая функция не может быть реализована. Однако соответствующее оптимальной решающей функции значение среднего риска $\rho(\delta_\theta|x, \theta)$ можно использовать для оценки качества неоптимальных решений, принимаемых при отсутствии полной информации о значении θ. Для любой решающей функции $\delta(z|x)$ потери вследствие неопределенности можно при каждом данном значении θ оценить разностью

$$\Delta(x, \theta) = \rho(\delta|x, \theta) - \rho(\delta_\theta|x, \theta),$$

а относительные потери — отношением $\Delta(x, \theta)/\rho(\delta_\theta|x, \theta)$.

Для оптимизации решений в условиях неопределенности применяются два основных подхода. Первый подход основан на том, что θ считается значением некоторой случайной величины Θ и риск заменяется его математическим ожиданием при данных x и z. Второй подход основан на том, что при каждой паре значений x, z риск заменяется его точной верхней гранью для всех возможных значений θ.

Предполагая, что θ представляет собой значение случайного вектора Θ, а входной сигнал x — значение случайного вектора X, зададим априорную плотность $\alpha(\theta)$ вектора Θ и условную плотность $p(x|\theta)$ вектора X при данном θ. Тогда апостериорная плотность $\omega(\theta|x)$ определится в соответствии с (4.23) и (4.8) формулой

$$\omega(\theta|x) = \frac{\alpha(\theta)\,p(x|\theta)}{\int \alpha(\eta)\,p(x|\eta)\,d\eta}. \tag{59}$$

Математическое ожидание риска при данных x и z определится формулой

$$r_1(z|x) = \int r(z|x, \theta)\,\omega(\theta|x)\,d\theta. \tag{60}$$

Заменив в (58) риск $r(x|z, \theta)$ его апостериорным математическим ожиданием $r_1(z|x)$, можно искать оптимальные или близкие к оптимальным решения, минимизирующие математическое ожидание $r_1(Z|x)$ при данном x, $\rho_1(\delta|x) = M[r_1(Z|x)|x]$.

Формула (59) представляет собой очевидное обобщение формулы Бейеса (1.23). Поэтому подход, основанный на замене риска его апостериорным математическим ожиданием, обычно называется *бейесовым*, а оптимальные решения в этом случае называются *бейесовыми решениями.*

Качество бейесова решения можно характеризовать апостериорным математическим ожиданием разности $\Delta(x|\theta) = \rho(\delta|x, \theta) - \rho(\delta_0|x, \theta)$:

$$\Delta_1(x) = \int \Delta(x, \theta)\,\omega(\theta|x)\,d\theta.$$

Слабостью бейесова подхода является необходимость задания априорного распределения $\alpha(\theta)$, для определения которого обычно нет никаких данных. Поэтому его приходится задавать произвольно, в частности, принято считать его равномерным в достаточно широкой области. Эта априорная неопределенность приводит к некоторой неопределенности в соответствующих оптимальных решениях. Однако, если размерность вектора x значительно больше размерности вектора θ (не меньше чем на $20 \div 30$), то апостериорная плотность $\omega(\theta|x)$ практически не зависит от априорной плотности $\alpha(\theta)$ и найденные изложенным способом оптимальные и близкие к оптимальным решения практически не зависят от $\alpha(\theta)$. Это в известной мере оправдывает бейесов подход. Поэтому он широко применяется на практике.

Желание усовершенствовать бейесов подход, освободить его от априорной неопределенности привело к появлению *эмпирического бейесова подхода*, основная идея которого состоит в том, чтобы путем проведения дополнительных опытов получить возможность оценивать $\omega(\theta|x)$ без знания $\alpha(\theta)$ [22, 88—91].

Если отказаться от предположения, что θ — возможное значение некоторой случайной величины, то естественно рассчитывать на наихудшее стечение обстоятельств. В этом случае при любых x и z риск $r(z|x, \theta)$ заменяется точной верхней гранью множества значений $r(z|x, \theta)$, соответствующих всем возможным значениям θ, т. е.

величиной

$$r_2(z \mid x) = \sup_\theta r(z \mid x, \theta).$$

Заменив в (58) риск $r(z \mid x, \theta)$ его верхней гранью $r_2(z \mid \theta)$, можно искать оптимальные и близкие к оптимальным решения, минимизирующие среднее значение $r_2(z \mid x)$ при данном x, $\rho_2(\delta \mid x) = M[r_2(Z \mid x) \mid x]$.

Оптимальные решения при таком подходе минимизируют максимально возможное значение риска. Поэтому подход, основанный на замене риска его верхней гранью, обычно называется *минимаксным*, а оптимальные решения в этом случае называются *минимаксными решениями*.

Качество минимаксного решения естественно характеризовать верхней гранью разности $\Delta(x, \theta) = \rho(\delta \mid x, \theta) - \rho(\delta_\theta \mid x, \theta)$:

$$\Delta_2(x) = \sup_\theta \Delta(x, \theta).$$

Достоинством минимаксного подхода является то, что он не требует задания каких-либо произвольных функций. Его недостаток состоит в том, что он основан на расчете на наихудшее стечение обстоятельств, которое, как правило, маловероятно. Поэтому минимаксные решения, минимизируя максимальные возможные потери, часто приводят к излишне большим потерям — малым выигрышам — при наиболее вероятных реальных стечениях обстоятельств. Иными словами, минимаксные решения чересчур осторожны.

5.6. Обучение моделей принятия решений. Если неопределенность слишком велика, т. е. разброс возможных значений риска при данном x слишком велик для любого решения z, то можно поставить вопрос о получении дополнительной информации для принятия решения с целью уменьшения неопределенности. Чем больше дополнительной информации получено, тем меньше будут потери. Однако получение любой информации связано с затратами средств и времени, что приводит к увеличению потерь. Таким образом, получение дополнительной информации приводит, с одной стороны, к уменьшению ожидаемых потерь за счет возможных последствий принимаемого решения, а с другой стороны, к увеличению потерь за счет стоимости информации. Поэтому возникает задача определения оптимальной дополнительной информации, при которой суммарные ожидаемые потери минимальны.

Пусть u — вектор, определяющий дополнительную информацию, $c(u|x)$ — затраты на получение информации u, приведенные к тем же единицам, в которых измеряется риск, т. е. потери, связанные с возможными последствиями принимаемого решения. Функция $c(u|x)$ в общем случае зависит от значения вектора u, от его размерности, характеризующей объем получаемой информации, и от того, какие именно величины служат его координатами.

При бейесовом подходе θ, x и u считаются возможными значениями некоторых случайных векторов Θ, X и U и для них задаются априорное распределение $\alpha(\theta)$ вектора Θ, условное распределение $p(x|\theta)$ вектора X при данном θ и условное распределение $q(u|x, \theta)$ вектора U при данных x и θ. Определив апостериорную плотность вероятности

$$\omega(\theta|x, u) = \frac{\alpha(\theta) p(x|\theta) q(u|x, \theta)}{\int \alpha(\eta) p(x|\eta) q(u|x, \eta) d\eta},$$

можно заменить в (58) риск его апостериорным математическим ожиданием

$$r_3(z|x, u) = \int r(z|x, \theta) \omega(\theta|x, u) d\theta$$

и после этого искать оптимальные или близкие к оптимальным решения, которые, конечно, будут зависеть от u. Поэтому обозначим оптимальную решающую функцию через $\delta_u(z|x)$. Суммарные потери для модели с такой решающей функцией при данных значениях x, θ и u определяются формулой

$$R(\delta_u|x, u, \theta) = \rho(\delta_u|x, u, \theta) + c(u|x) = \\ = \int r(z|x, \theta) \delta_u(z|x) dz + c(u|x). \quad (61)$$

Однако значение вектора u, так же как и значение параметра θ, заранее, до получения дополнительной информации, не известно. Поэтому для принятия решения о получении дополнительной информации, о ее составе и объеме необходимо определить средние потери, т. е. математическое ожидание потерь с учетом случайности величин Θ и U. Для этого следует умножить (61) на совместную плотность величин Θ и U при данном x и проинтегрировать по θ и u. В результате получим средние

суммарные потери при данном x:

$$\sigma_1(x) = \iint R(\delta_u \mid x,\ u,\ \theta)\, \omega(\theta \mid x)\, q(u \mid x,\ \theta)\, d\theta\, du. \quad (62)$$

В ряде задач решение о получении дополнительной информации приходится принимать до того, как будут известны исходные данные для принятия решения, т. е. вектор x. В таких случаях необходимо определить средние потери с учетом случайности всех трех величин Θ, U и X. Для этого следует умножить (61) на совместную плотность величин Θ, U и X и проинтегрировать по θ, u и x. Тогда получим средние суммарные потери

$$\sigma_1 = \iiint R(\delta_n \mid x,\ u,\ \theta)\, \alpha(\theta)\, p(x \mid \theta)\, q(u \mid x,\ \theta)\, d\theta\, dx\, du. \quad (63)$$

Задача определения оптимального объема и состава дополнительной информации сводится к минимизации величины (62) или (63) путем выбора состава (конкретного содержания) и размерности вектора u.

При минимаксном подходе получение дополнительной информации суживает пределы возможных значений θ и, следовательно, приводит к уменьшению верхней грани риска, т. е. к замене $r_2(z \mid x)$ величиной

$$r_4(z \mid x,\ u) = \sup_{\theta \in B(u)} r(z \mid x,\ \theta),$$

где $B(u)$ — область возможных значений θ при данном u. Получаемое в результате оптимальное решение тоже зависит от u. Поэтому, как и в предыдущем случае, обозначим оптимальную решающую функцию $\delta_u(z \mid x)$. Суммарные потери для модели с такой решающей функцией при данных значениях x, θ и u определяются формулой (61). Продолжая последовательно идею минимаксного подхода, будем вести расчет при неизвестных θ и u на наихудший случай, т. е. по максимальным суммарным потерям

$$\sigma_2(x) = \sup_{u} \sup_{\theta \in B(u)} R(\delta_u \mid x,\ u,\ \theta). \quad (64)$$

Если решение о получении дополнительной информации принимается до того, когда становится известным x, то величина $\sigma_2(x)$ должна быть заменена величиной

$$\sigma_2 = \sup_{x,\ u} \sup_{\theta \in B(u)} R(\delta_u \mid x,\ u,\ \theta). \quad (65)$$

Задача определения оптимального состава и объема информации в этом случае сводится к минимизации величины (64) или (65) путем выбора состава и размерности вектора u.

Улучшение качества решения путем ввода в модель дополнительной информации — вектора u — представляет собой *обучение* модели решения. Оптимизация процесса обучения достигается минимизацией одного из выражений (62) — (65) в зависимости от метода построения модели и условий ее применения (обучение до применения или обучение в процессе применения после наблюдения каждого очередного значения x). Бейесов подход к проблеме обучения автоматических систем развит в [75—80].

Кроме рассмотренных подходов к обучению модели принятия решений, возможны и другие подходы, основанные на различных методах оценивания неизвестных параметров. В связи с этим возможны и другие постановки и методы решения задачи оптимизации процесса обучения модели.

Кроме рассмотренных задач оптимизации без ограничений, в теории решений возникают задачи оптимизации решений при различных ограничениях. Ограничения могут накладываться на множество допустимых решений, как в примере 15, или на область допустимых значений z, или на какие-либо другие функции вида $s(z \mid x)$ (например, на другие координаты векторного риска). Методы решения таких задач также разрабатываются в современной теории решений.

ПРИЛОЖЕНИЯ

1. Импульсная дельта-функция и ее производные

Чтобы прийти к понятию импульсной δ-функции путем предельного перехода, рассмотрим функцию

$$\delta_l(x)=\begin{cases}\dfrac{1}{2l} & \text{при} \quad |x|<l,\\ 0 & \text{при} \quad |x|>l.\end{cases} \tag{1}$$

Эта функция называется в приложениях *единичным прямоугольным импульсом*.

Очевидно, что при любом $\varepsilon > l$

$$\int_0^{\varepsilon}\delta_l(x)\,dx=\int_{\varepsilon}^{0}\delta_l(x)\,dx=\frac{1}{2}. \tag{2}$$

Пусть $\varphi(x)$ — любая непрерывная функция. Вычислим интегралы

$$\int_a^b\varphi(x)\,\delta_l(u-x)\,dx, \qquad \int_a^b\varphi(x)\,\delta_l(x-u)\,dx.$$

На основании (1) при любом $u\in(a+l,\ b-l)$

$$\int_a^b\varphi(x)\,\delta_l(u-x)\,dx=\int_a^b\varphi(x)\,\delta_l(x-u)\,dx=\frac{1}{2l}\int_{u-l}^{u+l}\varphi(x)\,dx.$$

Применив к последнему интегралу теорему о среднем, получим

$$\int_a^b\varphi(x)\,\delta_l(u-x)\,dx=\int_a^b\varphi(x)\,\delta_l(x-u)\,dx=\varphi(x_0), \tag{3}$$

где $x_0\in(u-l,\ u+l)$. При любом $u\notin(a-l,\ b+l)$ оба интеграла равны нулю. Наконец,

$$\begin{aligned}\int_a^b\varphi(x)\,\delta_l(a-x)\,dx&=\int_a^b\varphi(x)\,\delta_l(x-a)\,dx=\frac{1}{2}\,\varphi(x_0'),\\ \int_a^b\varphi(x)\,\delta_l(b-x)\,dx&=\int_a^b\varphi(x)\,\delta_l(x-b)\,dx=\frac{1}{2}\,\varphi(x_0''),\end{aligned} \tag{4}$$

где $x_0'\in(a,\ a+l)$, $x_0''\in(b-l,\ b)$.

Перейдем теперь в предыдущих формулах к пределу при $l \to 0$. Ясно, что $\delta_l(x)$ при этом не имеет предела в обычном смысле. Однако все интегралы в (3) и (4) имеют вполне определенные пределы.

Рассуждая формально, приходим к выводу, что $\delta(x) = 0$ при любом $x \neq 0$, $\delta(0) = \infty$, причем площадь «всплеска» в бесконечность функции $\delta(x)$ в начале координат равна 1. Ясно, что $\delta(x)$ не существует как обычная функция. Она относится к классу обобщенных функций.

Формальный переход к пределу в (3) и (4) при $l \to 0$ дает

$$\int_a^b \varphi(x)\,\delta(u-x)\,dx = \int_a^b \varphi(x)\,\delta(x-u)\,dx = \varphi(u), \quad u \in (a, b), \tag{5}$$

$$\begin{aligned} \int_a^b \varphi(x)\,\delta(a-x)\,dx &= \int_a^b \varphi(x)\,\delta(x-a)\,dx = \frac{1}{2}\varphi(a), \\ \int_a^b \varphi(x)\,\delta(b-x)\,dx &= \int_a^b \varphi(x)\,\delta(x-b)\,dx = \frac{1}{2}\varphi(b). \end{aligned} \tag{6}$$

При $u \notin [a, b]$ оба интеграла в (5) равны 0.

Положив в первой формуле (6) $a = 0$, $b = \varepsilon$, а во второй $a = -\varepsilon$, $b = 0$, получим при $\varphi(x) \equiv 1$

$$\int_0^\varepsilon \delta(x)\,dx = \int_{-\varepsilon}^0 \delta(x)\,dx = \frac{1}{2} \quad \text{при любом} \quad \varepsilon > 0. \tag{7}$$

Функция $\delta(x)$ представляет собой предел функции $\delta_l(x)$ при $l \to 0$ в том смысле, что для любой непрерывной функции $\varphi(x)$

$$\int_{-\infty}^{\infty} \varphi(x)\,\delta(x)\,dx = \lim_{l \to 0} \int_{-\infty}^{\infty} \varphi(x)\,\delta_l(x)\,dx. \tag{8}$$

В таких случаях говорят, что $\delta_l(x)$ слабо сходится к $\delta(x)$ и $\delta(x)$ представляет собой слабый предел функции $\delta_l(x)$.

Таким образом, обобщенную функцию $\delta(x)$ можно определить совершенно строго как слабый предел обычной функции $\delta_l(x)$. Так определенная функция $\delta(x)$ называется *импульсной дельта-функцией* или, короче, *δ-функцией* *). Формулы (5), (6) и (7) представляют собой определяющие свойства δ-функции.

Заметим, что интегралы в (5)—(8) по определению представляют собой пределы соответствующих интегралов, содержащих функцию δ_l вместо δ-функции, при $l \to 0$. Только так следует понимать интегралы от обобщенных функций.

Из (5) и (6) при $\varphi(x) \equiv 1$ следует, что интеграл от δ-функции в пределах от $-\infty$ до u представляет собой единичную ступенча-

---

*) Дельта-функция была впервые введена в науку известным английским физиком Дираком. Поэтому ее часто называют дельта-функцией Дирака.

тую функцию:

$$\int_{-\infty}^{u} \delta(x)\,dx = 1(u) = \begin{cases} 0 & \text{при } u < 0, \\ 1/2 & \text{при } u = 0, \\ 1 & \text{при } u > 0. \end{cases} \qquad (9)$$

Отсюда следует, что δ-функция представляет собой производную единичной ступенчатой функции.

Совершенно так же, взяв четную функцию $\delta_l(x)$, имеющую непрерывные производные до порядка p включительно, определим производную δ-функции порядка p как слабый предел функции $\delta_l^{(p)}(x)$. При этом для любой функции $\varphi(x)$, непрерывной вместе со своими производными до порядка p включительно, будет справедлива формула

$$\int_{-\infty}^{\infty} \varphi(x)\,\delta^{(p)}(x)\,dx = (-1)^p \int_{-\infty}^{\infty} \varphi^{(p)}(x)\,\delta(x)\,dx = (-1)^p\,\varphi^{(p)}(0). \qquad (10)$$

Эту формулу можно принять за определение p-й производной δ-функции.

Из (10) следует, что для любого $u \in (a, b)$ и для любой функции $\varphi(x)$, непрерывной вместе со своими производными до порядка p включительно,

$$\int_{a}^{b} \varphi(x)\,\delta^{(p)}(u-x)\,dx = (-1)^p \int_{a}^{b} \varphi(x)\,\delta^{(p)}(x-u)\,dx = \varphi^{(p)}(u). \qquad (11)$$

Эти формулы получаются также из (5) формальным дифференцированием по u. Таким образом, в интегралах, содержащих δ-функцию или ее производные, можно выполнять дифференцирование по параметру под знаком интеграла сколько угодно раз.

Аналогично определяется δ-функция векторного аргумента. Очевидно, что δ-функцию векторного аргумента $x = \{x_1, \ldots, x_n\}$ можно определить как произведение δ-функций всех координат вектора x:

$$\delta(x) = \delta(x_1)\ldots\delta(x_n). \qquad (12)$$

Так как δ-функция в силу (7) обладает свойством (2.11) плотности, то ее можно рассматривать как слабый предел любой четной плотности при стягивании определяемого ею распределения в начало координат:

$$\delta(x) = \lim_{h\to\infty} hf(hx), \qquad (13)$$

где $f(y)$ — любая плотность.

В приложениях часто приходится пользоваться представлением δ-функции интегралом Фурье. Чтобы выразить δ-функцию интегралом Фурье, заметим, что любая непрерывная функция $\varphi(x)$, отличная от нуля только на некотором конечном интервале, может быть представлена интегралом Фурье:

$$\varphi(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda \int_{-\infty}^{\infty} \varphi(x)\,e^{i\lambda(x-u)}\,dx.$$

Изменив здесь формально порядок интегрирования, получим

$$\varphi(u)=\int_{-\infty}^{\infty}\varphi(x)\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\lambda(x-u)}\,d\lambda\right]dx.$$

Сравнив эту формулу с (5) при $a=-\infty$, $b=\infty$ и приняв во внимание, что обе формулы справедливы для любой функции $\varphi(x)$ рассматриваемого типа, приходим к выводу, что

$$\delta(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\lambda x}\,d\lambda. \tag{14}$$

Повторив эти рассуждения для случая векторного аргумента x, получим представление δ-функции n-мерного вектора x интегралом Фурье:

$$\delta(x)=\frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}e^{i\lambda^T x}\,d\lambda, \tag{15}$$

где векторы x и λ представлены в виде матриц-столбцов. Эта формула получается также из (14), если выразить δ-функцию вектора x формулой (12).

Интегралы в (14) и (15) расходящиеся. Ясно, что δ-функция не может выражаться сходящимся интегралом, так как любой сходящийся интеграл, зависящий от параметра, представляет собой обычную функцию этого параметра, в то время как δ-функция представляет собой обобщенную функцию.

2. Некоторые определенные интегралы

При изучении нормального распределения приходится пользоваться формулами

$$\int_{-\infty}^{\infty}e^{\eta t-\frac{1}{2}ct^2}\,dt=\sqrt{\frac{2\pi}{c}}\,e^{\frac{\eta^2}{2c}}, \tag{1}$$

$$\int_{-\infty}^{\infty}e^{\eta^T t-\frac{1}{2}t^T C t}\,dt=\sqrt{\frac{(2\pi)^n}{|C|}}\,e^{\frac{1}{2}\eta^T C^{-1}\eta}, \tag{2}$$

где η — в общем случае комплексная величина (скалярная в (1) и векторная в (2)), c — комплексная величина с положительной действительной частью, C — симметричная матрица с комплексными элементами, действительные части которых образуют строго положительно определенную матрицу, $|C|$ — определитель матрицы C.

Для вывода формулы (1) применим метод дифференцирования интеграла по параметру. Для этого положим

$$I(\eta)=\int_{\infty}^{\infty}e^{\eta t-\frac{1}{2}ct^2}\,dt.$$

Дифференцируя эту формулу по параметру η и интегрируя по частям, получаем

$$I'(\eta)=\int_{-\infty}^{\infty} te^{\eta t-\frac{1}{2}ct^2}\,dt=-\frac{1}{c}\int_{-\infty}^{\infty} e^{\eta t}\,de^{-\frac{1}{2}ct^2}=\frac{\eta}{c}\int_{-\infty}^{\infty} e^{\eta t-\frac{1}{2}ct^2}\,dt,$$

или

$$I'(\eta)=\frac{\eta}{c}\,I(\eta). \tag{3}$$

Таким образом, мы получили дифференциальное уравнение для $I(\eta)$. Чтобы полностью определить $I(\eta)$, необходимо добавить к уравнению (3) начальное условие. Для определения начального условия достаточно вычислить интеграл $I(\eta)$ для какого-нибудь одного значения параметра η. Проще всего вычислить

$$I(0)=\int_{-\infty}^{\infty} e^{-\frac{1}{2}ct^2}\,dt. \tag{4}$$

Так как этот интеграл не зависит от обозначения переменной интегрирования, то формулу (4) можно также написать в виде

$$I(0)=\int_{-\infty}^{\infty} e^{-\frac{1}{2}cs^2}\,ds. \tag{5}$$

Перемножая почленно равенства (4) и (5), получим

$$I^2(0)=\int_{-\infty}^{\infty} e^{-\frac{1}{2}ct^2}\,dt\int_{-\infty}^{\infty} e^{-\frac{1}{2}cs^2}\,ds=$$

$$=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} e^{-\frac{1}{2}cs^2}\,ds\right\}e^{-\frac{1}{2}ct^2}\,dt=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(t^2+s^2)}\,dt\,ds. \tag{6}$$

Рассмотрим сначала случай действительной (а следовательно, положительной) величины c. В этом случае двойной интеграл (6) легко вычисляется переходом к полярной системе координат. Положив

$$\sqrt{c}\,t=\rho\cos\varphi,\qquad \sqrt{c}\,s=\rho\sin\varphi$$

и имея в виду, что

$$J=\begin{vmatrix}\frac{\partial t}{\partial\rho} & \frac{\partial t}{\partial\varphi}\\ \frac{\partial s}{\partial\rho} & \frac{\partial s}{\partial\varphi}\end{vmatrix}=\begin{vmatrix}\frac{1}{\sqrt{c}}\cos\varphi & -\frac{\rho}{\sqrt{c}}\sin\varphi\\ \frac{1}{\sqrt{c}}\sin\varphi & \frac{\rho}{\sqrt{c}}\cos\varphi\end{vmatrix}=\frac{\rho}{c},$$

получим

$$I^2(0)=\frac{1}{c}\int_0^{2\pi} d\varphi\int_0^{\infty} e^{-\frac{1}{2}\rho^2}\rho\,d\rho=\frac{2\pi}{c}.$$

Отсюда находим

$$I(0)=\int_{-\infty}^{\infty} e^{-\frac{1}{2}ct^2}\,dt=\sqrt{\frac{2\pi}{c}}. \tag{7}$$

Рассмотрим теперь случай комплексной величины c. Положив $c=|c|e^{i\varphi}$, $t=e^{-i\varphi/2}\tau$, приведем интеграл (4) к виду

$$I(0)=e^{-\frac{i\varphi}{2}}\int_{-\infty e^{i\varphi/2}}^{\infty e^{i\varphi/2}} e^{-\frac{1}{2}|c|\tau^2}\,d\tau, \tag{8}$$

где интегрирование производится по прямой комплексной плоскости, проходящей через начало координат под углом $\varphi/2$ к действительной оси. Для вычисления полученного интеграла рассмотрим на комплексной плоскости переменной τ замкнутый контур, состоящий из отрезка действительной оси $(0, R)$, дуги окружности $(R, Re^{i\varphi/2})$ радиуса R и отрезка $(Re^{i\varphi/2}, 0)$ прямой, по которой производится интегрирование в (8) (рис. 38). Так как подынтегральная функция является аналитической функцией, не имеющей особых точек в области, ограниченной этим контуром, то интеграл по этому контуру равен 0 при любом R. А так как $\mathrm{Re}\{c\}>0$ и, следовательно, $|\varphi|<\pi/2$ и $\mathrm{Re}\{\tau^2\}>0$, то интеграл по дуге окружности $(R, Re^{i\varphi/2})$ стремится к 0 при $R\to\infty$. Поэтому при $R\to\infty$ получаем равенство

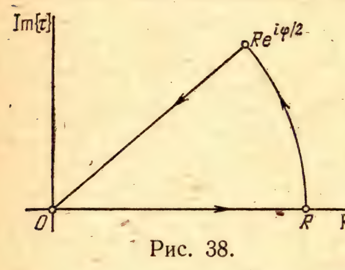


Рис. 38.

$$\int_0^{\infty} e^{-\frac{1}{2}|c|\tau^2}\,d\tau+\int_{\infty e^{i\varphi/2}}^{0} e^{-\frac{1}{2}|c|\tau^2}\,d\tau=0,$$

откуда следует, что

$$\int_0^{\infty e^{i\varphi/2}} e^{-\frac{1}{2}|c|\tau^2}\,d\tau=\int_0^{\infty} e^{-\frac{1}{2}|c|\tau^2}\,d\tau.$$

Аналогично приходим к выводу, что

$$\int_{-\infty e^{i\varphi/2}}^{0} e^{-\frac{1}{2}|c|\tau^2}\,d\tau=\int_{-\infty}^{0} e^{-\frac{1}{2}|c|\tau^2}\,d\tau.$$

Следовательно,

$$\int_{-\infty e^{i\varphi/2}}^{\infty e^{i\varphi/2}} e^{-\frac{1}{2}|c|\tau^2}\,d\tau=\int_{-\infty}^{\infty} e^{-\frac{1}{2}|c|\tau^2}\,d\tau.$$

Подставив это выражение в (8) и пользуясь формулой (7), получаем

$$I(0)=e^{-\frac{i\varphi}{2}}\sqrt{\frac{2\pi}{|c|}}=\sqrt{\frac{2\pi}{|c|\,e^{i\varphi}}}=\sqrt{\frac{2\pi}{c}}.$$

Таким образом, формула (7) справедлива и для любого комплексного c с положительной действительной частью.

Интегрируя дифференциальное уравнение (3) при начальном условии (7), получаем

$$I(\eta)=\sqrt{\frac{2\pi}{c}}\,e^{\frac{\eta^2}{2c}},$$

что и доказывает формулу (1).

Для доказательства формулы (2) приведем матрицу C ортогональным преобразованием к диагональной форме. В результате получим $A^TCA=\Lambda$, где Λ — диагональная матрица, элементами которой служат собственные значения $\lambda_1, \ldots, \lambda_n$ матрицы C (n — размерность вектора t в (2) и порядок матрицы C), а A — ортогональная матрица, $A^TA=I$. Преобразование $\eta=A\xi$, $t=As$ дает $\eta^Tt=\xi^TA^TAs=\xi^Ts$, $t^TCt=s^TA^TCAs=s^T\Lambda s$, и интеграл (2) приводится к виду (с учетом равенства $|A|=1$)

$$\int_{-\infty}^{\infty} e^{\eta^Tt-\frac{1}{2}t^TCt}\,dt=\int_{-\infty}^{\infty} e^{\xi^Ts-\frac{1}{2}s^T\Lambda s}\,ds=\prod_{k=1}^{n}\int_{-\infty}^{\infty} e^{\xi_ks_k-\frac{1}{2}\lambda_ks_k^2}\,ds_k.$$

Таким образом, интеграл (2) приводится к произведению интегралов (1). Пользуясь формулой (1), находим

$$\int_{-\infty}^{\infty} e^{\eta^Tt-\frac{1}{2}t^TCt}\,dt=\sqrt{\frac{(2\pi)^n}{\lambda_1\ldots\lambda_n}}\,e^{\frac{1}{2}\xi^T\Lambda^{-1}\xi}.$$

Подставив сюда выражение $\xi=A^{-1}\eta=A^T\eta$ и имея в виду, что произведение собственных значений матрицы равно ее определителю, $\lambda_1\ldots\lambda_n=|C|$, и что $A\Lambda^{-1}A^T=(A\Lambda A^T)^{-1}=C^{-1}$, получаем формулу (2).

3. Таблицы

Таблица I

Функция Лапласа $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int\limits_0^u e^{-t^2/2}\, dt$

	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2703	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3437	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	47725	47778	47831	47882	47932	47981	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48746	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49135	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49597	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3	0,49865	49903	49931	49952	49966	49977	49984	49989	49993	49995
4,0	0,499968									
4,5	0,499997									
5,0	0,49999997									

Таблица 2

Производные функции Лапласа

Функция $\Phi'(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3639	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3760	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблица 2 (*продолжение*)

Производные функции Лапласа

u	$\Phi'(u)$	$\Phi''(u)$	$\Phi'''(u)$	$\Phi^{(IV)}(u)$	$\Phi^{(V)}(u)$	$\Phi^{(VI)}(u)$	$\Phi^{(VII)}(u)$
0,0	+0,3989	0,0000	−0,3989	0,0000	+1,1968	0,0000	−5,9841
0,1	+0,3970	−0,0397	−0,3930	+0,1187	+1,1671	−0,5915	−5,7762
0,2	+0,3910	−0,0782	−0,3754	+0,2315	+1,0799	−1,1420	−5,1711
0,3	+0,3814	−0,1144	−0,3471	+0,3330	+0,9413	−1,6142	−4,2223
0,4	+0,3683	−0,1473	−0,3094	+0,4184	+0,7607	−1,9777	−3,0122
0,5	+0,3521	−0,1760	−0,2641	+0,4841	+0,5501	−2,2114	−1,6448
0,6	+0,3332	−0,1999	−0,2133	+0,5278	+0,3231	−2,3052	−0,2324
0,7	+0,3122	−0,2186	−0,1593	+0,5486	+0,0937	−2,2601	+1,1136
0,8	+0,2897	−0,2318	−0,1043	+0,5469	−0,1247	−2,0880	+2,2938
0,9	+0,2661	−0,2395	−0,0506	+0,5244	−0,3203	−1,8095	+3,2303
1,0	+0,2420	−0,2420	0,0000	+0,4839	−0,4839	−1,4518	+3,8715
1,1	+0,2179	−0,2396	+0,0458	+0,4290	−0,6091	−1,0458	+4,1959
1,2	+0,1942	−0,2330	+0,0854	+0,3635	−0,6926	−0,6230	+4,2103
1,3	+0,1714	−0,2228	+0,1182	+0,2918	−0,7341	−0,2130	+3,9475
1,4	+0,1497	−0,2096	+0,1437	+0,2180	−0,7364	+0,1590	+3,4595
1,5	+0,1295	−0,1943	+0,1619	+0,1457	−0,7042	+0,4736	+2,8109
1,6	+0,1109	−0,1775	+0,1730	+0,0781	−0,6440	+0,7181	+2,0712
1,7	+0,0940	−0,1599	+0,1777	+0,0176	−0,5632	+0,8870	+1,3078
1,8	+0,0790	−0,1421	+0,1769	−0,0341	−0,4692	+0,9809	+0,5801
1,9	+0,0656	−0,1247	+0,1713	−0,0760	−0,3693	+1,0058	−0,0647
2,0	+0,0540	−0,1080	+0,1620	−0,1080	−0,2700	+0,9718	−0,5939
2,1	+0,0440	−0,0924	+0,1500	−0,1302	−0,1765	+0,8915	−0,9899
2,2	+0,0355	−0,0780	+0,1362	−0,1436	−0,0927	+0,7784	−1,2488
2,3	+0,0283	−0,0652	+0,1215	−0,1492	−0,0214	+0,6460	−1,3788
2,4	+0,0224	−0,0538	+0,1066	−0,1483	+0,0362	+0,5064	−1,3965
2,5	+0,0175	−0,0438	+0,0920	−0,1424	+0,0800	+0,3697	−1,3242
2,6	+0,0136	−0,0353	+0,0782	−0,1328	+0,1105	+0,2438	−1,1864
2,7	+0,0104	−0,0281	+0,0656	−0,1207	+0,1293	+0,1338	−1,0076
2,8	+0,0079	−0,0222	+0,0541	−0,1073	+0,1379	+0,0429	−0,8097
2,9	+0,0060	−0,0173	+0,0441	−0,0934	+0,1385	−0,0281	−0,6110
3,0	+0,0044	−0,0133	+0,0354	−0,0798	+0,1330	−0,0798	−0,4254
3,1	+0,0033	−0,0101	+0,0281	−0,0669	+0,1231	−0,1140	−0,2624
3,2	+0,0024	−0,0076	+0,0220	−0,0552	+0,1107	−0,1332	−0,1271
3,3	+0,0017	−0,0057	+0,0170	−0,0444	+0,0969	−0,1404	−0,0213
3,4	+0,0012	−0,0042	+0,0130	−0,0359	+0,0829	−0,1384	+0,0561
3,5	+0,0009	−0,0030	+0,0098	−0,0282	+0,0694	−0,1300	+0,1078
3,6	+0,0006	−0,0022	+0,0073	−0,0219	+0,0570	−0,1176	+0,1380
3,7	+0,0004	−0,0016	+0,0054	−0,0168	+0,0460	−0,1030	+0,1510
3,8	+0,0003	−0,0011	+0,0039	−0,0127	+0,0365	−0,0878	+0,1512
3,9	+0,0002	−0,0008	+0,0028	−0,0095	+0,0284	−0,0730	+0,1426
4,0	+0,0001	−0,0005	+0,0020	−0,0070	+0,0218	−0,0594	+0,1286

Таблица 3

Двусторонние границы *T*-распределения: значения t_α, определяемые уравнением $\int_{-t_\alpha}^{t_\alpha} s_k(t)\,dt = \alpha$

k \ α	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	142	289	445	617	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	137	277	424	584	765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	134	271	414	569	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	132	267	408	559	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	131	265	404	553	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	130	263	402	549	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	130	262	399	546	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	129	261	398	543	703	883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	129	260	397	542	700	879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	129	260	396	540	697	876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	128	259	395	539	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	128	259	394	538	694	870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	128	258	393	537	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	128	258	393	536	691	866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	128	258	392	535	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	128	257	392	534	689	863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	127	257	392	534	688	962	1,067	1,330	1,734	2,103	2,552	2,878	3,922

Таблица 3 *(продолжение)*

k \ α	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
19	127	257	391	533	688	861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	127	257	391	533	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	127	257	391	532	686	859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	127	256	390	532	686	858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	127	256	390	532	685	858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	127	256	390	531	685	857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	127	256	390	531	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	127	256	390	531	684	856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	127	256	389	531	684	855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	127	256	389	530	683	855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	127	256	389	530	683	854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	127	256	389	530	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	126	255	388	529	681	851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	126	254	387	527	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	126	254	386	526	677	845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	126	253	385	524	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Таблица 4

Двусторонние границы χ^2-распределения: значения ε_α, определяемые уравнением

$$\int\limits_{k/(1+\varepsilon_\alpha)^2}^{k/(1-\varepsilon_\alpha)^2_+} p_k(z)\,dz = \alpha \text{ *)}$$

k \ α	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,568	906	1,602	2,946	6,923				
2	367	473	0,678	1,125	2,086	3,400	5,857	8,500	
3	290	370	482	0,730	1,270	1,932	3,000	4,200	9,00
4	248	306	398	563	0,941	1,382	2,056	2,700	5,00
5	221	277	348	475	738	1,104	1,594	2,000	3,80
6	200	251	308	416	623	0,918	1,306	1,650	3,00
7	185	232	290	380	576	800	1,143	1,393	2,50
8	173	216	269	354	516	713	0,986	1,225	2,05
9	162	202	252	329	476	650	889	1,094	1,75
10	153	192	239	304	442	596	814	0,980	1,50
12	140	176	218	276	388	527	700	840	1,30
14	130	162	290	252	357	468	620	740	1,14
16	122	150	188	236	325	422	564	671	1,02
18	115	143	177	223	297	390	500	600	0,92
20	108	136	168	210	282	370	480	567	85
25	096	122	148	187	247	317	408	485	70
30	088	111	137	172	226	281	369	425	60
35	085	101	127	156	207	261	347	400	56
40	076	095	119	146	193	242	312	375	52
45	071	089	112	139	184	228	288	350	48
50	068	084	105	133	174	212	270	311	45
60	062	077	095	122	155	193	242	283	40
70	057	072	088	112	145	180	222	250	37
80	054	067	082	103	138	167	200	236	35
90	051	063	078	096	131	151	192	220	32
100	048	060	074	092	125	146	184	200	30
150	040	050	060	075	096	125	146	167	0,225
200	034	042	053	065	084	100	133	144	190
250	031	038	048	058	076	091	115	135	175
500	022	028	031	041	054	064	077	085	125
1000	016	019	025	037	044	047	056	059	080

*) $(1-\varepsilon_\alpha)_+ = \max(1-\varepsilon_\alpha,\ 0)$.

Таблица 5

Верхние 100 (1—α)%-ные точки χ^2-распределения: значения χ^2_α, определяемые уравнением $P_k(\chi^2_\alpha)=\int\limits_0^{\chi^2_\alpha} p_k(z)\,dz=\alpha$

k \ α	0,01	0,02	0,05	0,10	0,20	0,30	0,50	0,70	0,80	0,90	0,95	0,98	0,99	0,999
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,71	3,84	5,41	6,64	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	0,115	0,185	0,352	0,584	1,005	1,424	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	0,554	0,752	1,145	1,610	2,34	3,00	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	0,872	1,134	1,635	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	1,239	1,564	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,02	14,07	16,62	18,48	24,3
8	1,646	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,78	13,44	15,99	18,31	21,2	23,2	29,6
11	3,05	3,61	4,58	5,58	6,99	8,15	10,34	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	3,57	4,18	5,23	6,30	7,81	9,03	11,34	14,01	15,81	18,55	21,0	24,1	26,2	32,9
13	4,11	4,76	5,89	7,04	8,63	9,93	12,34	15,12	16,98	19,81	22,4	25,5	27,7	34,5
14	4,66	5,37	6,57	7,79	9,47	10,82	13,34	16,22	18,15	21,1	23,7	26,9	29,1	36,1
15	5,23	5,98	7,26	8,55	10,31	11,72	14,34	17,32	19,31	22,3	25,0	28,3	30,6	37,7
16	5,81	6,61	7,96	9,31	11,15	12,62	15,34	18,42	20,5	23,5	26,3	29,6	32,0	39,3
17	6,41	7,26	8,67	10,08	12,00	13,53	16,34	19,51	21,6	24,8	27,6	31,0	33,4	40,8
18	7,02	7,91	9,39	10,86	12,86	14,44	17,34	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	7,63	8,57	10,12	11,65	13,72	15,35	18,34	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	8,26	9,24	10,85	12,44	14,58	16,27	19,34	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	8,90	9,92	11,59	13,24	15,44	17,18	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	9,54	10,60	12,34	14,04	16,31	18,10	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	10,20	11,29	13,09	14,85	17,19	19,02	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	10,86	11,99	13,85	15,66	18,06	19,94	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	11,52	12,70	14,61	16,47	18,94	20,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
26	12,20	13,41	15,38	17,29	19,82	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	12,88	14,12	16,15	18,11	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	13,56	14,85	16,93	18,94	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	14,26	15,57	17,71	19,77	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	14,95	16,31	18,49	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9	59,7

Таблица 6*)

Верхние 5%-ные и 1%-ные точки F-распределения: значения f_α, определяемые уравнением

$$F_{lk}(f_\alpha) = \int_0^{f_\alpha} f_{lk}(f)\,df = \alpha$$

Значения $f_{0,95}$ (верхние 5%-ные точки)

k	l									
	1	2	3	4	5	6	8	12	24	∞
1	161,40	199,50	215,70	224,60	230,20	234,00	238,90	243,90	249,00	254,30
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6.16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2.85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,33	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

*) Эта таблица заимствована из [35].

Значения $f_{0,99}$ (верхние 1%-ные точки)

	l									
	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,66	2,34	1,95	1,38
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,00

Таблица 7

Предельное распределение Колмогорова $K(u)=\sum_{\nu=-\infty}^{\infty}(-1)^{\nu}e^{-2\nu^2u^2}$

	0	1	2	3	4	5	6	7	8	9
0,3	0,0000	0000	0000	0001	0002	0003	0005	0008	0013	0019
0,4	0028	0040	0055	0074	0097	0126	0160	0200	0247	0300
0,5	0361	0428	0503	0585	0675	0772	0876	0987	1104	1228
0,6	1357	1492	1632	1778	1927	2080	2236	2396	2558	2722
0,7	2888	3055	3223	3391	3560	3728	3896	4064	4230	4395
0,8	4559	4720	4880	5038	5194	5347	5497	5645	5791	5933
0,9	6073	6209	6343	6473	6601	6725	6846	6964	7079	7191
1,0	7300	7406	7508	7608	7704	7798	7889	7976	8061	8143
1,1	8223	8300	8374	8445	8514	8580	8644	8706	8765	8823
1,2	8878	8930	8981	9030	9076	9121	9164	9206	9245	9283
1,3	9319	9354	9387	9418	9449	9478	9505	9531	9556	9580
1,4	9603	9625	9646	9665	9684	9702	9718	9734	9750	9764
1,5	9778	9791	9803	9815	9826	9836	9846	9855	9864	9873
1,6	9880	9888	9895	9902	9908	9914	9919	9924	9929	9934
1,7	9938	9942	9946	9950	9953	9956	9959	9962	9965	9967
1,8	9969	9971	9973	9975	9977	9979	9980	9981	9983	9984
1,9	9985	9986	9987	9988	9989	9990	9991	9991	9992	9993
2	99933	99970	99987	99995	99998	99999	1,00000			

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. Андерсон Т. В. (Anderson T. W.). An introduction to multivariate statistical analysis.— John Wiley, 1958. Русский перевод: Введение в многомерный статистический анализ.—М.: Физматгиз, 1963.
2. Бартлетт М. С. (Bartlett M. S.). On the theory of statistical regression.— Proc. Roy. Soc. Edinburgh, 1933, v. 53, 260—283.
3. Бартлетт М. С. (Bartlett M. S.). An introduction to stochastic processes with special reference to methods and applications.— Cambridge University Press, 1955. Русский перевод: Введение в теорию случайных процессов.—М.: ИЛ, 1958.
4. Блекуэлл Д., Гиршик М. А. (Blackwell D., Girshick M. A.). Theory of games and statistical decisions.— John Wiley, 1954. Русский перевод: Теория игр и статистических решений. — М.: ИЛ, 1958.
5. Блюм Дж. Р. (Blum J. R.) Approximation methods which converge with probability one.—Ann. Math. Stat., 1954, v. 25, 382—386.
6. Блюм Дж. Р. (Blum J. R.). Multidimensional stochastic approximation method. — Ann. Math. Stat., 1954, v. 25, 737—744.
7. Бокс Дж. Е. П. (Box G. E. P.). Multifactor designs of first order.— Biometrica, 1952, v. 39, 49—57.
8. Бокс Дж. Е. П., Уилсон К. Б. (Box G. E. P., Wilson K. B.). An experimental attainment of optimum conditions.— Journ. Roy. Stat. Soc., Ser. B, 1951, v. 13, 1—45.
9. Бокс Дж. Е. П., Хантер Дж. С. (Box G. E. P., Hunter J. S.). Multifactor experimental designs for exploring response surfaces. — Ann. Math. Stat., 1957, v. 28, 195—241.
10. Бородачев Н. А. Основные вопросы теории точности производства. — Изд. АН СССР, 1950.
11. Бусленко Н. П., Голенко Д. И., Соболь И. М., Сраговин В. Г., Шрейдер Ю. А. Метод статистических испытаний (метод Монте-Карло).—М.: Физматгиз, 1962.
12. Бутон Р. К. (Booton R. C.). The analysis of nonlinear control systems with random inputs.—Proc. Sympos. Nonlinear Circuit Analysis, 1953, v. 2, 369—391.
13. Вазан М. Т. (Wasan M. T.). Stochastic approximation.— Cambridge University Press, 1969. Русский перевод: Стохастическая аппроксимация.—М.: Мир, 1972.
14. Вальд А. (Wald A.) Sequential analysis.— John Wiley, 1947. Русский перевод: Последовательный анализ.—М.: Физматгиз, 1960.
15. Вальд А. (Wald A.). Statistical decision functions.— John Wiley, 1950. Русский перевод: приложение «Статистические решающие функции» к книге «Позиционные игры» (ред. Н. Н. Воробьев, Н. И. Врублевская). — М.: Наука, 1967,

16. Гантмахер Ф. Р. Теория матриц.— М.: Наука, 1966.
17. Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин.— М.: Гостехиздат, 1949.
18. Головина Л. И. Линейная алгебра и некоторые ее приложения.— М.: Наука, 1975.
19. Госсет В. С. (Gosset W. S., «Student»). The probable error of a mean.— Biometrika, 1908, v. 6, 1—25.
20. Дашевский М. Л. Уравнения семиинвариантов нелинейной динамической системы.— Автом. и телемех., 1968, № 10, 63—71.
21. Де Гроот М. (De Groot M. H.) Optimal statistical decisions.— McGraw-Hill, 1970. Русский перевод: Оптимальные статистические решения.— М.: Мир, 1974.
22. Джонс Г. В. (Johns H. V.) Non-parametric empirical Bayes procedures.— Ann. Math. Stat., 1957, v. 28, № 4, 649—669.
23. Добровидов А. В. Об одном алгоритме непараметрической оценки случайных многомерных сигналов.— Автом. и телемех., 1971, № 2, 88—89.
24. Дрейпер Н., Смит Г. (Draper N., Smith H.) Applied regression analysis.— John Wiley, 1966. Русский перевод: Прикладной регрессионный анализ.— М.: Статистика, 1973.
25. Дуда Р., Харт П. (Duda R. O., Hart P. E.). Pattern classification and scene analysis.— John Wiley 1973. Русский перевод: Распознавание образов и анализ сцен.— М.: Мир, 1976.
26. Дэвид Ф. Н. (David F. N.). A note on unbiased limits for the correlation coefficient.— Biometrika, 1937, v. 29, 157—160.
27. Дэвид Ф. Н. (David F. N.). Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples.— Cambridge University Press, 1938.
28. Дюге Д. (Dugue D.) Application des proprietés de la limite au sens du calcul des probabilités à l'étude des diverses questions d'estimation.— Journ. Ecol. Polytechn., 1937, v. 3, № 4, 305—372.
29. Емельянов С. В., Озерной В. М. Decision making in multi-objective problems: a survey.— Probl. Contr. Inform. Theory, 1975, Jubilee number, 51—64.
30. Ермаков С. М. Метод Монте-Карло и смежные вопросы.— М.: Наука., 1975.
31. Казаков И. Е. Приближенный вероятностный анализ точности работы существенно нелинейных систем.— Автом. и телемех., 1956, т. 17, № 5, 385—409.
32. Казаков И. Е. Статистический анализ систем с многомерными нелинейностями.— Автом. и телемех., 1965, № 3, 463—469.
33. Калман Р. Е. (Kalman R. E.). A new approach to linear filtering and prediction problems.— Trans. ASME, Journ Basic Engin., 1960. v. 82D, 35—45.
34. Кашьяп Р. Л., Рао А. Р. (Kashyap R. L., Rao A. R.). Dynamic stochastic models from empirical data.— Academic Press, 1976.
35. Кендалл М. Дж., Стьюарт А. (Kendall M. J., Stewart A.). The advanced theory of statistics. Charles Griffin, v. 1. Distribution theory, 1958; v. 2, Inference and relationship, 1961; v. 3, Design and analysis and time-series, 1966. Русский перевод: Теория распределений.— М.: Наука, 1966; Статистические

выводы и связи.— М.: Наука, 1973; Многомерный статистический анализ и временные ряды.— М.: Наука, 1976.
36. Кифер Дж., Вольфовиц Дж. (Kiefer J., Wolfowitz J.). Stochastic estimation of the maximum of a regression function.— Ann. Math. Stat., 1952, v. 23, 462—466.
37. Клоппер К. Дж., Пирсон Е. С. (Clopper С. J., Pearson Е. S.). The use of confidence or fiducial limits illustrated in the case of the binomial. — Biometrika, 1934, v. 26, 404—413.
38. Кокрен В. Дж., Кокс Дж. М. (Cochran W. G., Cox G. M.). Experimental designs.— John Wiley, 1957.
39. Колмогоров А. Н. Общая теория меры и теория вероятностей.— Сборник трудов секции точных наук Коммунистической академии, 1929, т. 1, 8—21.
40. Колмогоров А. Н. Grundbegriffe der Wahrscheinlichkeitsrechnung. — Springer-Verlag, 1933. Русский перевод: Основные понятия теории вероятностей.— М.: ОНТИ, 1936;—М.: Наука, 1974.
41. Колмогоров А. Н. Sulla determinazione empirica di una legge di distribuzione.— Giorn. Inst. Ital. Attuari, 1933, v. 4, 83—91.
42. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа.— М.: Наука, 1972.
43. Крамер Г. (Cramer Н.). On the composition of elementary errors.— Skand. Aktuan, 1928, v. 11, 13—74, 141—180.
44. Крамер Г. (Cramer Н.). Random variables and probability distributions.—Cambridge Tracts in Mathematics, № 36, Cambridge University Press, 1937. Русский перевод: Случайные величины и распределения вероятностей.— М.: ИЛ, 1947.
45. Крамер Г. (Cramer Н.). Mathematical methods of statistics.— Princeton University Press, 1946. Русский перевод: Математические методы статистики.— М.: ИЛ, 1948;— М.: Наука, 1975.
46. Красовский Н. Н. Некоторые задачи теории устойчивости движения.— М.: Физматгиз, 1959.
47. Купер Д. В. (Cooper D. V.). Adaptive pattern recognition and signal detection using stochastic approximation.— IEEE Trans. on Electronic Computers, 1964, EC-13, № 3, 306—307.
48. Ланкастер П. (Lankaster P.). Theory of matrices.—Academic Press, 1969. Русский перевод: Теория матриц.— М.: Наука, 1978.
49. Леман Э. Л. (Lehmann E. L.). Testing statistical hypotheses.— John Wiley. 1959. Русский перевод: Проверка статистических гипотез.— М.: Наука, 1964.
50. Лоули Д. Н., Максвелл А. Е. (Lawley D. N., Maxwell A. E.). Factor analysis as a statistical method.— Butterworths, London, 1963. Русский перевод: Факторный анализ как статистический метод.— М.: Мир, 1967.
51. Лоэв М. (Loève М.). Probability theory. — D. Van Nostrand, 1960. Русский перевод: Теория вероятностей.— М.: ИЛ, 1962.
52. Льюс Р. Д., Райфа Г. (Luce R. D., Raiffa H.). Games and decisions. Introduction and critical survey. — John Wiley, 1957. Русский перевод: Игры и решения. Введение и критический обзор.— М.: ИЛ, 1961.
53. Ляпунов А. М. Общая задача об устойчивости движения.— Харьковск. матем. общество, 1892; — М.: Гостехиздат, 1950.

54. Ляпунов А. М. Sur une proposition de la théorie des probabilités.— Bull. Acad. Sciences de St. Peterbourg, 1900, v. 13, 359—386.
55. Ляпунов А. М. Nouvelle forme de théorème sur la limite de probabilité.— Mem. Acad. Sciences de St. Peterbourg, 1901, v. 12, 1—24.
56. Мальцев А. И. Основы линейной алгебры.— М.: Наука, 1975.
57. Маркус М., Минц Х. (Marcus M., Minc H.). A survey of matrix theory and matrix inequalities.— Allyn and Bacon, 1964. Русский перевод: Обзор по теории матриц и матричных неравенств.— М.: Наука, 1972.
58. Масси Ф. Дж. (Massey F. J.). A note on the estimation of distribution function by confidence limits.— Ann. Math. Stat., 1950, v. 21, 116—119.
59. Mathematical methods for digital computers (ed. Ralfston A., Wilf H. S.).— John Wiley, 1962.
60. Мизес Р. (von Mises R.). Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.— Leipzig—Wien, 1931.
61. Моррис У. Т. (Morris W. T.). Management science. A Bayesian introduction.— Prentice-Hall, 1968. Русский перевод: Наука об управлении. Бейесовский подход.— М.: Мир, 1971.
62. Надарая Э. А. О непараметрических оценках плотности вероятности и регрессии.—Теор. вер. и ее прим., 1965, т. 10, вып. 1, 190—194.
63. Налимов В. В., Чернова О. А. Планирование экспериментов.— М.: Наука, 1971.
64. Невё Ж. (Neveu J.). Bases mathématiques du calcul des probabilités.— Masson, 1964. Русский перевод: Математические основы теории вероятностей.— М.: Мир, 1969.
65. Нейман Дж., Пирсон Е. С. (Neyman J., Pearson E. S.). On the use and interpretation of certain test criteria for purposes of statistical inference.— Biometrika, 1928, v. 20 A, Part I, 175—240, Part II, 263—294.
66. Нейман Дж., Пирсон Е. С. (Neyman J., Pearson E. S.). On the problem of the most efficient tests of statistical hypotheses.— Trans. Roy. Soc. London, Ser. A, 1933, v. 231, 289—337.
67. Нейман Дж., Пирсон Е. С. (Neyman J., Pearson E. S.). Contributions to the theory of testing statistical hypotheses.— Stat. Research Mem., 1936, v. 1, 1—37; 1938, v. 2, 25—57.
68. Нейман Дж., Пирсон Е. С. (Neyman J., Pearson E. S.). Sufficient statistics and uniformly most powerful tests of statistical hypotheses.— Stat. Research Mem., 1936, v. 1, 113—137.
69. Никольский С. М. Курс математического анализа, т. 1, т. 2. —М.: Наука, 1973.
70. Парзен Е. (Parzen E.). On estimation of a probability density function and mode.— Ann. Math. Stat., 1962, v. 33, 1065—1076.
71. Пирсон К. (Pearson K.). Contributions to the mathematical theory of evolution.— Phil. Trans. Roy. Soc., 1894, v. 185, 71—121; 1895, v. 186, 343—365.
72. Пирсон К. (Pearson K.). On the criterion that a given system of devations from the probable in the case of a correlated system

of variables in such that it can be reasonably supposed to have arisen from random sampling.—Phil. Mag., 1900, v. 50, 157—175.
73. Пирсон К. (Pearson K.) On lines and planes of closest fit to systems of points in space.— Phil. Mag., 1901, v. 2, 559—572.
74. Пирсон К. (Pearson K.) On the systematic fitting of curves to observations and measurements.— Biometrika, 1902, v. 1, 265—278; v. 2, 1—27.
75. Пугачев В. С. Статистические проблемы теории распознавания образов. Труды III Всесоюзн. совещ. по автом. управлению (Одесса, 20—25 сент. 1965). Самонастраивающиеся системы. Распознавание образов.— М.: Наука, 1967, 192—202.
76. Пугачев В. С. Оптимальные алгоритмы обучения автоматических систем в случае неидеального учителя.—Докл. АН СССР, 1967, т. 172, № 5, 1039—1042.
77. Пугачев В. С. Оптимальные обучающиеся системы.— Докл. АН СССР, 1967, т. 175, № 5, 1022—1025.
78. Пугачев В. С. Оптимальное обучение автоматических систем в изменяющихся условиях.— Автом. и телемех., 1967, № 10, 153—167.
79. Пугачев В. С. Статистическая теория обучающихся автоматических систем.—Изв. АН СССР, Тех. киб., 1967, № 6, 26—42.
80. Пугачев В. С. Optimal learning systems.— Kybernetika, 1971, v. 7, № 1, 347—376.
81. Пугачев В. С. Рекуррентное оценивание переменных и параметров в стохастических системах, описываемых разностными уравнениями. — Докл. АН СССР, 1978, т. 243, № 5, 1131—1133.
82. Пугачев В. С. Оценивание переменных и параметров в дискретных нелинейных системах. — Автом. и телемех., 1979, № 4.
83. Райфа Г. (Raiffa H.). Decision analysis. Addison-Wesley, 1969. Русский перевод: Анализ решений.— М.: Наука, 1977.
84. Рао С. Р. (Rao C. R.). Information and accuracy attainable in the estimation of statistical parameters.—Bull. Calcutta Math. Soc., 1945, v. 37, 81—91.
85. Рао С. Р. (Rao C. R.). Sufficient statistics and minimum variance estimates.—Proc. Cambr. Phil. Soc., 1949, v. 45, 213—218.
86. Рао С. Р. (Rao C. R.). Linear statistical inference and its applications.— John Wiley, 1965. Русский перевод: Линейные статистические методы и их применения.— М.: Наука, 1968.
87. Растригин Л. А. Статистические методы поиска.— М.: Наука, 1968.
88. Роббинс Г. (Robbins H.). An empirical Bayes approach to statistics.—Proc. 3d Berkeley Symp. on math. stat. and prob., 1955, v. 1, 157—164.
89. Роббинс Г. (Robbins H.). An empirical Bayes approach to nonparametric two-way classification studies in item analysis and prediction (ed. H. Solomon).— Stanford University Press, 1961, 221—232.
90. Роббинс Г. (Robbins H.). An empirical Bayes approach to testing statistical hypotheses.— Review of the Internat. Stat. Inst., 1963, v. 31, № 2, 195—208.
91. Роббинс Г. (Robbins H.). An empirical Bayes approach to problems of statistical decision theory.—Ann. Math. Stat., 1964, v. 35, № 1, 1—20.

92. Роббинс Г., Монро С. (Robbins H., Monro S.). A stochastic approximation method.— Ann. Math. Stat., 1951, v. 22, 400—407.
93. Розенблат М. А. (Rosenblatt M. A.). Remarks on some nonparametric estimates of a density function.—Ann. Math. Stat., 1956, v. 27, 642—669.
94. Сборник научных программ на Фортране. System/360 scientific subroutine pakage (360 A-CM-03X) version III. Programmer's Manual. IBM, Technical Publications Department, 1960—1971. Русский перевод: Под ред. С. Я. Виленкина: вып. 1, Статистика; вып. 2, Матричная алгебра и линейная алгебра.— М.: Статистика, 1975.
95. Синицын И. Н. Методы статистической линеаризации (обзор).— Автом. и телемех., 1974, № 5, 74—94.
96. Смирнов В. И. Курс высшей математики.—М.: Наука, т. 1; т. 2; т. 3, ч. 1, ч. 2; т. 4, ч. 1, 1974; т. 4, ч. 2; т. 5, 1975.
97. Смирнов Н. В. Sur la distribution de ω^2.— Comptes Rendus Acad. Sci. de Paris, 1936, v. 202, 449—453.
98. Смирнов Н. В. Оценка расхождения между кривыми распределения в двух независимых выборках.— Бюлл. Моск. университета, 1930, т. 2, № 2, 3—14.
99. Снедекор Г. В. (Snedecor G. W.). Statistical methods. Ames.— Iowa State College Press, 1940.
100. Соболь И. М. Численные методы Монте-Карло. —М.: Наука, 1973.
101. Сомервилл М. Дж., Атертон Д. П. (Somerville M. J., Atherton D. P.). Multi-gain representation for a single-valued nonlinearity with several inputs, and the evaluation of the equivalent gains by a cursor method.— Proc. IEE, Part C, Monograph № 309M, 1958, 1—13.
102. Уайльд Д. (Wilde D.). Optimum seeking method.—John Wiley, 1964. Русский перевод: Методы поиска экстремума.— М.: Физматгиз, 1967.
103. Уилкс С. (Wilks S.). Mathematical Statistics.—Princeton University Press, 1944. Русский перевод: Математическая статистика.— М.: Наука, 1967.
104. Уишарт Дж. (Wishart J.). The generalized product moment distribution in samples from a normal multivariate population.— Biometrika, 1928, v. 20 A, 353—361.
105. Уишарт Дж., Бартлетт М. С. (Wishart J., Bartlett M. S.). The distribution of the second order moment statistics in a normal system.— Proc. Cambridge Phil. Soc., 1932, v. 28, 455—459.
106. Фабиан В. (Fabian V.). Стохастическая аппроксимация для отыскания оптимальных условий при обработке результатов экспериментов и в самонастраивающихся системах.— Aplikace mathematiky, 1961, v. 6, 162—183 (на чешск. яз.).
107. Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры.—М.: Физматгиз, 1963.
108. Финни Д. Дж. (Finney D. J.). An introduction to the theory of experimental design.— University of Chicago Press, 1960. Русский перевод: Введение в теорию планирования экспериментов.—М.: Наука, 1970.
109. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления.—М.. Физматгиз, т. 1, 1958; т. 2, 1959; т. 3, 1960.

110. Фишер Р. А. (Fisher R. A.). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population.— Biometrika, 1915, v. 10, 507—521.
111. Фишер Р. А. (Fisher R. A.). On the mathematical foundations of theoretical statistics.— Phil. Trans. Roy. Soc., ser. A, 1921, v. 222, 309—368.
112. Фишер Р. А. (Fisher R. A.). On a distribution yielding the error functions of several well-known statistics.—Proc. Int. Math. Congress, Toronto, 1924, v. 2, 805—813.
113. Фишер Р. А. (Fisher R. A.). Statistical methods for researsh workers.— Oliver and Boyd, lt ed., 1925, 12th ed., 1954. Русский перевод: Статистические методы для исследователей.—М.: Госстатиздат, 1958.
114. Фишер Р. А. (Fisher R. A.). Theory of statistical estimation.— Proc. Cambridge Phil. Soc., 1925, v. 22, 700—725.
115. Фишер Р. А. (Fisher R. A.). Applications of «Student's» distribution.— Metron, 1926, v. 5, № 4, 90—104.
116. Фишер Р. А. (Fisher R. A.). The design of experiments.— Oliver and Boyd, 1935.
117. Фишер Р. А., Иэйтс Ф. А. (Fisher R. A., Yates F. A.). Statistical tables for biological, agricultural and medical research. — Oliver and Boyd, 1957.
118. Харман Г. Г. (Harman H. H.). Modern factor analysis.—University of Chicago Press, 1960. Русский перевод: Современный факторный анализ.— М.: Статистика, 1972.
119. Хеннекен П. Л., Тортра А. (Hennequin P. L., Tortrat A.). Théorie des probabilités et quelques ses applications.— Dunod, 1970. Русский перевод: Теория вероятностей и некоторые ее приложения.— М.: Наука, 1974.
120. Хинчин А. Я. Асимптотические законы теории вероятностей.— ОНТИ, 1936.
121. Цыпкин Я. З. Адаптация и обучение в автоматических системах.— М.: Наука, 1968.
122. Цыпкин Я. З. Основы теории обучающихся систем.— М.: Наука, 1970.
123. Шеффе Г. (Scheffe H.). The analysis of variance. John Wiley, 1959. Русский перевод: Дисперсионный анализ.— М.: Физматгиз, 1963.
124. Ширяев А. Н. Статистический последовательный анализ.— М.: Наука, 1976.
125. Шустер Е. Ф. (Schuster E. F.). Estimation of probability density function and its derivatives.— Ann. Math. Stat., 1969, v. 40, 1187—1195.
126. Эджуорт Ф. И. (Edgeworth F. Y.). The law of error.— Proc. Cambridge Phil. Soc., 1905, v. 20, 36—65.

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

$A, B, C, \ldots$ — события, множества, в некоторых случаях матрицы

$\bigcup\limits_{s \in S} A_s$ или $\sum\limits_{s \in S} A_s$ — объединение (сумма) событий, множеств

$\bigcap\limits_{s \in S} A_s$ — пересечение (произведение) событий, множеств

$B \subset A$ — событие (или множество) B, представляет собой подсобытие (подмножество) события (множества) A

$\bar{A}$ — событие, противоположное A

ω — элементарное событие

Ω — достоверное событие, пространство элементарных событий

$\varnothing$ — невозможное событие

$\mathcal{S}$ — поле событий множеств, алгебра событий, σ-алгебра событий

$A \in \mathcal{S}$ — событие принадлежит полю событий, множеств

$X, Y, Z, \ldots$ — случайные величины, скалярные или векторные

$x, y, z, \ldots$ — возможные значения или реализации случайных величин $X, Y, Z, \ldots$

$f(x)$ — плотность случайной величины X

$F(x)$ — функция распределения случайной величины X

$f(x, y)$ — совместная плотность случайных величин X, Y

$F(x, y)$ — совместная функция распределения случайных величин X, Y

$f(x \mid y)$ — условная плотность случайной величины X при данном значении y случайной величины Y

$F(x \mid y)$ — условная функция распределения случайной величины X при данном значении y случайной величины Y

$\mathbf{1}(x)$ — единичная функция, равная 1 при $x > 0$, 1/2 при $x = 0$ и 0 при $x < 0$

$\delta(x)$ — дельта-функция Дирака, производная от $\mathbf{1}(x)$

$\mathbf{1}_B(x)$ — индикатор области B, т. е. функция, равная 1 в области B и 0 вне B

$v(B)$ — мера области B (длина, площадь, объем, ...)

d_x^* — выборочная дисперсия скалярной случайной величины X

σ_x^* — выборочное среднее квадратическое отклонение случайной величины X

k_{xy}^* — выборочная ковариация случайных величин X и Y

r_{xy}^* — выборочный коэффициент корреляции случайных величин X и Y

$m_x = MX$ — математическое ожидание (среднее значение) случайной величины X

$X^0 = X - m_x$ — центрированная случайная величина X

$D_x = DX$ — дисперсия случайной величины X

k_{xy} — ковариация случайных величин X и Y

r_{xy} — коэффициент корреляции случайных величин X и Y

γ_x — момент второго порядка (второй момент) скалярной случайной величины X

γ_{xy} — смешанный момент второго порядка случайных величин X и Y

$\gamma_x(a)$ — момент второго порядка случайной величины X относительно точки a

$\gamma_{xy}(a, b)$ — смешанный момент второго порядка случайных величин X и Y относительно точек a и b

Γ_x — момент второго порядка (второй момент) векторной случайной величины X, оператор момента второго порядка

K_x — ковариационная матрица случайного вектора X, ковариационный оператор

R_x — корреляционная матрица случайного вектора X

Γ_{xy} — взаимный момент второго порядка (второй момент) случайных векторов X и Y, взаимный оператор момента второго порядка

α_r — момент порядка r скалярной случайной величины

μ_r — центральный момент порядка r скалярной случайной величины

$\alpha_{k_1 \dots k_n}$ — момент порядка r случайного вектора $(k_1 + \dots + k_n = r)$

$\mu_{k_1 \dots k_n}$ — центральный момент порядка r случайного вектора $(k_1 + \dots + k_n = r)$

γ_1 — асимметрия или скошенность распределения случайной величины

γ_2 — эксцесс распределения случайной величины

$\zeta_{1/2}$ — медиана скалярной случайной величины

ζ_p — квантиль порядка p скалярной случайной величины

$\zeta_{1/4}$, $\zeta_{3/4}$ — квартили

$E = (\zeta_{3/4} - \zeta_{1/4})/2$ — срединное или вероятное отклонение или семиинтерквартильная широта

$\zeta_{0,1}, \dots, \zeta_{0,9}$ — децили

$\Phi(u)$ — функция Лапласа

$M[Y \mid x]$ — условное математическое ожидание случайной величины Y при данном значении x случайной величины X

$g(\lambda)$ — характеристическая функция случайной величины

$g_1 * g_2$ — свертка g_1 и g_2

$\varkappa_r$ — семиинвариант (кумулянт) порядка r скалярной случайной величины

$\varkappa_{h_1 \dots h_n}$ — семиинвариант порядка r векторной случайной величины $(h_1 + \dots + h_n = r)$

$P_{m,n}$ — вероятность того, что событие появится m раз при n независимых опытах

$\varphi_n(u)$ — производящая функция для вероятностей $P_{m,n}$

$R_{k,n}$ — вероятность того, что при n независимых опытах событие появится не меньше k раз

$\lambda(t)$ — интенсивность пуассоновского потока

μ — среднее число событий, происходящих на данном временном интервале, параметр распределения Пуассона

$p_k(z)$ — плотность χ^2-распределения с k степенями свободы

T — стьюдентово отношение, обобщенное стьюдентово отношение

$s_k(t)$ — плотность распределения Стьюдента (T-распределения) с k степенями свободы

$w_{mn}(s)$ — плотность распределения Уишарта

$f_{lk}(f)$ — плотность F-распределения с l и k степенями свободы

$N(m, K)$ — нормальное распределение с параметрами m и K

$K(u)$ — предельное распределение Колмогорова

$H_n(x)$ — полиномы Эрмита степени n

$\hat{\theta}$ — оценка статистической характеристики θ

$e(\varphi)$ — эффективность оценки $\hat{\Theta} = \varphi(U)$

$g(u \mid \theta)$ — функция правдоподобия

$o(x)$ — величина высшего порядка малости по сравнению с величиной x

$O(x)$ — величина одинакового порядка малости по сравнению с величиной x

$\operatorname{tr} A$ — след матрицы A

$\mathrm{B}(p, q)$ — бета-функция

$\Gamma(z)$ — гамма-функция

T — верхний индекс — символ транспонирования

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

1 р. 20 к.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА